Experimentally-based Mathematical Modeling to Analyze T Helper 17 Cell Differentiation in Heterogeneous Cell Populations

Yat Hin CHAN

MSc Thesis
December 2015

UNIVERSITY OF HELSINKI
Faculty of Science
Department of Mathematics and Statistics

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Mathematics and Statistics

Tekijä — Författare — Author
Yat Hin CHAN

Työn nimi — Arbetets titel — Title
Experimentally-based Mathematical Modeling to Analyze T Helper 17 Cell Differentiation in Heterogeneous Cell Populations

Oppiaine — Läroämne — Subject
Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	December 2015	76

Tiivistelmä — Referat — Abstract

During the recent years, there has been an increasing interest among both biologists and mathematicians to model and understand gene regulatory mechanisms that drive cell differentiation processes. Mathematical modeling of these processes is often based on the assumption of homogeneous cell population. However, in many applications the cell populations of interest can be heterogeneous. For example, $CD4^+$ T cell populations that are studied in this thesis may consist of many distinct T helper (Th) cell subtypes. Consequently, cell populations in cell differentiation studies are inevitably heterogeneous.

In this thesis, we develop a new modeling approach that takes the possibility of a heterogeneous population into account and apply this approach to study the Th17 cell differentiation. More specifically, we design ordinary differential equation (ODE) models that take the heterogeneity into account by describing approximative subpopulations that evolve in parallel within a population and have cell type specific regulatory mechanisms and dynamics. In our application, we allow the cell population to be split into two subpopulations, an activated T helper (Th0) cell subpopulation and an actively differentiating Th17 cell subpopulation. Both Th0 and Th17 cell dynamics share the same rate parameters to describe the common reaction mechanisms within the subtypes. Three models, homogeneous population (M_1), replicate-independent heterogeneous population (M_2) and replicate-dependent heterogeneous population (M_3), are constructed.

In order to infer Th17 cell differentiation dynamics and to detect possible heterogeneity during differentiation in a data-driven manner, we combine mathematical modeling with RNA sequencing (RNA-Seq) data using statistical modeling. To carry out posterior analysis, we use Bayesian inference with population-based Markov chain Monte Carlo (popMCMC) sampling method. Our results show strong evidence for the replicate-dependent heterogeneous population model (M_3) evolving in Th17 lineage polarizing condition. In addition, the model makes it possible to predict the resulting molecular dynamics.

Avainsanat — Nyckelord — Keywords
Mathematical Modeling, ODE, Bayesian, MCMC, Cell Differentiation, Heterogeneous Population.

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpula Science Library

Muita tietoja — Övriga uppgifter — Additional information

Preface

This master thesis has been carried out in the Computational Systems Biology group at Aalto University, Espoo, Finland. I am deeply grateful to my supervisors, Dr. Jukka Intosalmi and Prof. Harri Lähdesmäki, for provided me with this memorable research opportunity and giving many suggestions and comments in this project. I also want to acknowledge Prof. Jukka Corander, for supervising, reading and examining the thesis.

Further, I am obliged to Prof. Riitta Lahesmaa and the members of the Molecular Systems Immunology group at University of Turku and Åbo Akademi, Turku, Finland, for carrying out the RNA sequencing data and Sini Rautio, M.Sc., for carrying out the data alignment and pre-processing analysis. I am also indebted to Kari Nousiainen, M.Sc., for inspiring me to use the form of matrix differential equation with analytic solution and Mr. Henrik Mannerström for suggesting me some mathematical and computational ideas.

In addition, I am very happy to work with the whole research group and to live with my friends in Finland. The most important thing is that I wish to thank my family for supporting me to study and work aboard.

Acronyms

Biology-related

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
tRNA	transfer RNA
Tc cell	cytotoxic T cell / killer T cell / $CD8^+$ T cell
Th cell	T helper cell / $CD4^+$ T cell
Thp cell	T helper precursor cell / naïve $CD4^+$ T cell
Th0 cell	activated T helper cell
Th17 cell	T helper 17 cell
iTreg cell	induced regulatory T cell
TCR	T cell receptor
TF	transcription factor
STAT	signal transducer and activator of transcription
RORC	retinoic acid receptor-related orphan receptor C
RORγt	retinoic acid receptor-related orphan receptor gamma thymus
FOXP3	forkhead box P3
T-bet	T-box 21 / TBX21
GATA3	trans-acting T-cell-specific transcription factor
TGF-β	transforming growth factor-beta
IFN	interferon
IL	interleukin
GRN	gene regulatory network
RNA-Seq	RNA sequencing
NGS	next-generation sequencing
RC	read count
RPKM	reads per kilobase per million

Mathematics-related

ODE	ordinary differential equation
MDE	matrix differential equation
LLN	law of large numbers

Bayesian-related

MCMC	Markov chain Monte Carlo
i.i.d.	independent and identically distributed
MH	Metropolis-Hastings
AM	adaptive Metropolis
popMCMC	population-based MCMC
PHM	posterior harmonic mean
PAM	prior arithmetic mean

Population Models

M_1	homogeneous population model
M_2	replicate-independent heterogeneous population model
M_3	replicate-dependent heterogeneous population model

Contents

Chapter 1

Introduction

In this thesis, we develop a mathematical modeling approach that can be used to approximately detect the heterogeneity in a cell population. Since the standard way of modeling to assume homogeneous population does not hold in many applications, we solve this issue by our new modeling approach, which takes a heterogeneous population into account. Then we use this new approach to study a real application, the T helper 17 (Th17) cell differentiation in a data-driven manner using statistical models.

T helper (Th) cells, which are a subtype of lymphocytes, mediate adaptive immunity to provide defense against infection and there are four well characterized lineages of Th cells, Th1, Th2, Th17 and induced regulatory T (iTreg) cells, having their own functions [1]. Th17 cells play a critical role in the adaptive immune system and induce signature cytokines, interleukin (IL)-17A, IL-17F and IL-22 to promote inflammatory, which is important in clearing pathogens during host defense reactions and inducing tissue inflammation in autoimmune diseases [2, 3, 4]. It is well-known that Th cell differentiation is driven by a particular cytokine environment. Given T cell receptor (TCR) activation by antigens in the presence of cytokine signals, Th cells may differentiate into its lineages [1]. The differentiation process from a naïve $CD4^+$ T (Thp) cell into a Th17 cell can be induced by cytokines, IL-6 and transforming growth factor (TGF)-β, and mainly regulated by transcription factors (TFs), retinoic acid receptor-related orphan receptor gamma thymus (RORγt) and signal transducer and activator of transcription 3 (STAT3) [2].

Although the Th lineage is well-known and mathematical modeling has been applied to biological systems for decades, the differentiation dynamics is poorly understood and difficult to be determined. There are several mathematical models

demonstrating the dynamical networks of Th cell differentiation during the recent years [5, 6, 7, 8, 9, 10, 11]. Most of the studies construct mechanistic models using homogeneous population [5, 6, 7, 10, 11]. Schulz et al. [5] have introduced a positive feedback loops model, which can explain expression kinetics of T-bet, IL-12Rβ2, and interferon (IFN)-γ during Th1 cell priming. Höfer et al. [7] have created a mathematical model describing the regulatory mechanisms of trans-acting T-cell-specific transcription factor (GATA3) activity in Th2 cells with the bistable expression levels. Mendoza [10] has presented the core regulatory network of a discrete dynamical system controlling the differentiation of Th cells, reconstructed from published molecular data. Mendoza et al. [11] have presented a continuous dynamical system to model the signaling network that controls the differentiation process of Th cells from Thp cell to all four well-known subtypes, Th1, Th2, Th17 and iTreg. On the other hand, some of them study differentiation dynamics that takes the heterogeneity into account in modeling but focuses on the stability analysis of the steady-state behavior [8, 9]. Hong et al. [8] have proposed a mathematical model for the reciprocal differentiation of Th17 and iTreg cells with a generic form of ordinary differential equations (ODEs) showing that antigen-activated Thp cells treated with TGF-β with cell-to-cell variability can differentiate into a heterogeneous population of effector cells with distinct phenotypes such as forkhead box P3 (FOXP3)-only, RORγt-only and double-expressing cells. Hong et al. [9] have also presented a simple theoretical framework to show how heterogeneous differentiation in a two-master-regulator paradigm can be governed by a signaling network motif common to all subtypes of Th cells.

The main goal of this thesis is to develop a data-driven modeling approach that can be applied to approximately study gene regulatory networks (GRNs) and cell differentiation dynamics. The mathematical models are formulated by using parallel ODE systems describing the dynamics of the specific cell types. In cell differentiation processes, it is possible that only a fraction of cells gets activated and therefore constructs a heterogeneous cell population. In the application, we consider heterogeneity and apply this new approach on the Th17 cell differentiation. In Th17 polarizing conditions, we assume that a fixed fraction of cells starts to actively go through the differentiation process becoming Th17 cells and rest of the cells remain activated only as activated Th (Th0) cells. Our modeling approach takes two parallel ODE systems that approximately describe the dynamics of GRNs of Th0 and Th17 cells. The systems illustrate expressed genes dynamics and describe reactions including transcription, translation and degradation of TFs, STAT3 and RORC, in both messenger RNA (mRNA) and protein levels. All the reactions are assumed to be linear that allows the systems to be written in the form of matrix differential equations (MDEs). Therefore, analytical explicit solutions are available to present model responses. Two network systems are similar and share the

common reaction rates. The differences between the cell type specific dynamics is that the activation signal applied on Th17 cells is boosted by the cytokine inputs and there is a feedback transcription self-regulated by RORC in Th17 cells. In order to compare homogeneous and heterogeneous population structures, three alternative models including one model in homogeneous population (M_1) and two models in heterogeneous population (M_2, M_3), are generated. Since experiments are repeated with replicates, population proportions can be replicate-independent (M_2) or replicate-dependent (M_3). The former one shares the same population proportion between replicates and the latter one has an individual proportion for each replicate.

Furthermore, we compare alternative models using statistical modeling with experiential data, which is RNA sequencing (RNA-Seq) measurements from humans. The treatment of the discrete read count data is based on the negative binomial distribution that is commonly used in numerous biological applications [12, 13, 14, 15, 16]. We then use Bayesian inference that provides a reliable framework to handle the data and draw conclusions represented by posterior probability distributions based on Bayes' rule [17]. To carry out posterior analysis, we use population-based Markov chain Monte Carlo (popMCMC) sampling to estimate target distributions [18]. Then we can estimate marginal likelihood of each model through the power posterior estimator based on thermodynamic integration [18, 19, 20] to rank alternative models that produce similar dynamics but cannot be selected based on visual inspection. The result shows that the replicate-dependent heterogeneous population model (M_3) is the decisively top-ranked model with the strongest evidence (i.e. the highest marginal likelihood) in the human RNA-Seq data.

The content of this thesis is the following that links between molecular biology, mathematical modeling and Bayesian analysis. In Chapter 2, introduction to molecular biology is provided for readers to understand biological background. There are general cellular components and processes, the role of Th cell lineages and Th cell differentiation, and also RNA-Seq. Chapter 3 covers the topics of mathematical modeling and we present how mathematical models are used in biology to describe dynamical biochemical systems. Simple ODE examples on reaction networks and the general solution of MDE, which is a special case of ODEs, are included in Chapter 3. Chapter 4 and 5 cover Bayesian analysis with computational steps. Chapter 4 introduces the ideas of Bayesian inference and statistical models, likelihood functions and prior distributions. Chapter 4 also covers the analysis of model ranking and posterior predictions. Chapter 5 shows the computational methods of MCMC sampling algorithms and the marginal likelihood estimator, the power posterior estimator. Chapter 6 presents the main idea of the

thesis, which is the heterogeneous population structured modeling approach. It also presents the application on Th cell biology, including the Th cell networks and alternative models. The computational implementation is included at the end of Chapter 6. Chapter 7 shows the final results of this project including parameter estimation and model ranking of three alternative models from both simulated and experimental data. Posterior predictions from the experimental data are provided by the top-ranked model, M_3. Chapter 8 is about the conclusions and the future plans on FOXP3 dynamics of iTreg cells. The last part, Chapter 9 is the appendix including the probability distributions used in likelihood and prior and other figures obtained from this study.

Chapter 2

Biological Background

In this chapter, we present basic information on molecular biology for readers to understand the biological application background of the whole thesis. It includes the general cellular components and processes following the book [21]. It also includes the role of T helper (Th) cell lineages and Th cell differentiation [1, 2, 3, 4, 5, 7, 8, 9]. At the end of this chapter, we introduce how the experimental data is obtained by the experimental technology, RNA sequencing (RNA-Seq) [13, 14, 15, 16, 21, 22].

2.1 Cellular Components and Processes

The cell is the basic structural unit of living organisms and the following introduces cellular components and processes that can be found in the book [21]. The human cell except red blood cell contains nucleus, which is the control center of the cell and stores most of the genetic material. The cell contains many biomolecules such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), transcription factors (TFs) and genes, which are all important components in the cellular processes. DNA is a molecule that stores hereditary information within a cell and carries this information from generation to generation. RNA is a molecule similar to DNA, but in single-stranded instead of double-stranded as DNA. There are several types of RNA, in which messenger RNA (mRNA) carries information from DNA to ribosome for translation. TF is a protein that binds with DNA controlling the transcription rate of genetic information from DNA to mRNA. A gene is a DNA segment representing a piece of genetic code that consists of exons and introns. Both of them exist in an initial RNA transcript known as pre-mRNA but introns are removed and exons are joined to form mature mRNA after RNA splicing.

Therefore, exons code for proteins but introns do not. Figure 2.1 shows that a gene is a DNA segment consisting of exons and introns.

Each cell type has the same DNA sequence but its own pattern of regulated gene expression. Gene expression is the process by which genetic codes are used in synthesis of gene products. A gene is a combination of expressed and repressed genes along a DNA sequence. Expressed genes are used in the synthesis of gene products, often proteins. There are many steps in the gene expression mechanism regulated by TFs in both mRNA and protein levels, for example, transcription, RNA processing, RNA transport and localization, translation, mRNA degradation and protein transport. In this thesis, we focus only on the transcription and translation processes, as shown in Figure 2.1. Transcription is the synthesis of mRNA from a DNA template in the nucleus, whereas translation is the synthesis of proteins from an mRNA template in the cytoplasm. While transcription process, TF binds to and moves along DNA to produce mRNA. While translation process, ribosome binds to mRNA and then transfer RNA (tRNA) carries methionine to ribosome while ribosome moves along mRNA to produce proteins.

Cell differentiation, which is the main cellular process in this thesis, is the process of changing cell type from one to another to become more specialized. A human body is composed of trillions of cells and each cell type has its own specialized function. Cell differentiation involves a switch from one gene expression pattern to another. Therefore, regulation of gene expression including transcription and translation is distinct between cell types. In fact, there are many other cellular processes such as cell division. Cell division, which refers to mitosis instead of meiosis in our application, is the process of making new body cells. Mitosis is a one-step process dividing into two identical daughter cells. Before a parent cell dividing, two identical DNA copies are produced from a DNA molecule and this process is called DNA replication.

2.2 T Helper Cells

In this section, we introduce several lineages of T helper (Th) cells and their roles in adaptive immune system following the studies [1, 2, 3, 4, 5, 7, 8, 9]. Further, we present the well-known cell differentiation discoveries of each Th cell type found in the studies [1, 5, 7, 8, 9].

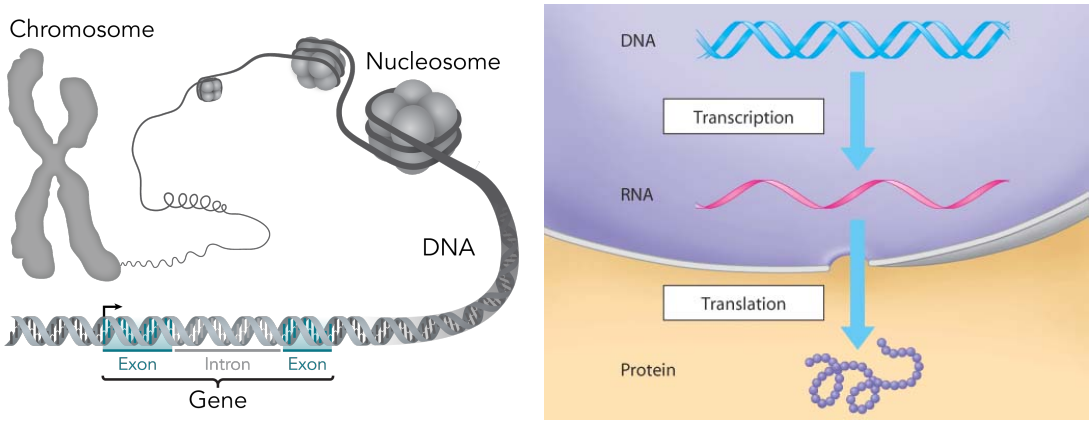

Figure 2.1: The left figure shows the structure of a chromosome, which is formed by DNA. A gene consisting of exons and introns is a segment of DNA. The right figure shows transcription and translation that produces RNA from DNA and protein from RNA, respectively.

http://upload.wikimedia.org/wikipedia/commons/e/e1/Chromosome-DNA-gene.png

http://wps.pearsoncustom.com/wps/media/objects/4557/4666819/ebook/htm/chp10_6.htm

2.2.1 Roles of T Helper Cells

T cells play an important role in the adaptive immune system. The immune system is a system of lymphocytes and other cell types within an organism to provide defense against infection. There are two types of immune system in vertebrates and they are innate and adaptive. T cells, as known as T lymphocytes, are a type of lymphocyte responsible for T-cell-mediated adaptive immune response. Figure 2.2 shows that T cells include cytotoxic T (Tc) cells and T helper (Th) cells, and actually regulatory T cells that are not shown in the figure. In fact, Th cells do not kill infected cells as Tc cells, but do activation of Tc cells instead. Th cells also help B cells as known as B lymphocytes to produce antibodies, enhance the innate immune system and suppress or regulate the immune response to avoid autoimmunity [1].

There are several lineages of Th cells discovered by biologists [1], mainly including Th1, Th2, Th17 and induced regulatory T (iTreg) cells, as shown in Figure 2.3 and Table 2.1. They are all important and play main roles to response the immune system by inducing signature cytokines. Hence, they have their own dynamics during cell differentiation. Th1 cells mainly induce interferon(IFN)-γ as signature cytokine to be the host immunity effectors against intracellular bacteria and protozoa [5], whereas interleukin(IL)-4, IL-5 and IL-13 are induced by Th2 cells to against extracellular parasites including helminths [7]. Th17 cells are pro-inflammatory

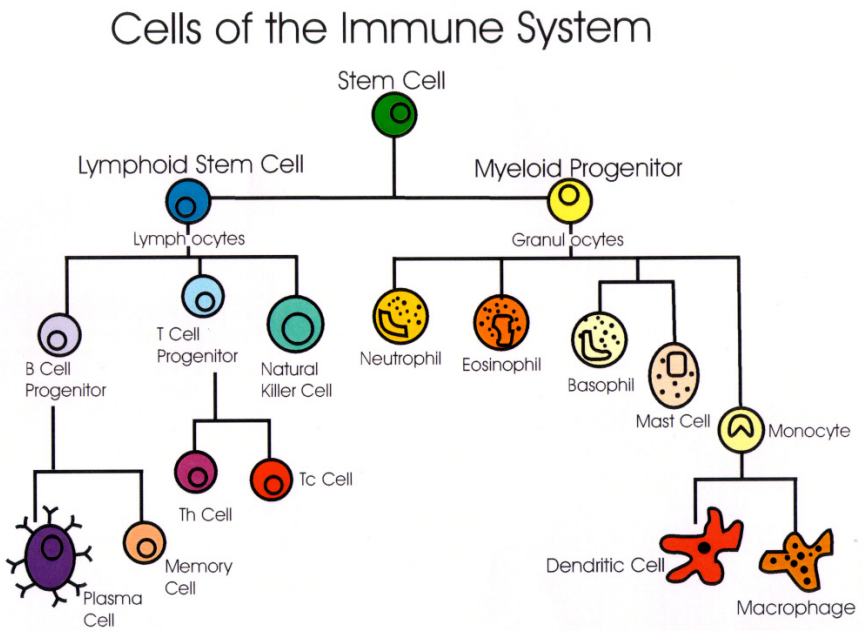

Figure 2.2: Different cell types of the immune system. The cytotoxic T (Tc) cells and T helper (Th) cells are subtypes of T cells, whereas T cells are a subtype of lymphocytes.

http://www.biologymad.com/StudentsWork/12%20-%20immunologynotes.pdf

by inducing signature cytokines, IL-17A, IL-17F and IL-22, to promote inflammatory, whereas iTreg cells are anti-inflammatory by inducing IL-10 and transforming growth factor(TGF)-β to reduce inflammatory response [8, 9]. Th17 cells are important in clearing pathogens during host defense reactions and inducing tissue inflammation in autoimmune diseases [2, 3, 4]. iTreg cells have suppressive function and maintain immunological tolerance in autoimmune disease. Table 2.1 lists the lineages with their corresponding signature cytokines.

2.2.2 T Helper Cell Differentiation

T Helper (Th) cell differentiation is the development process of Th cells that requires an activation under a particular cytokine environment and is mainly regulated by its key transcription factor (TF), as known as master regulator. Given T cell receptor (TCR) activation by antigens, naïve $CD4^+$ T (Thp) cells may differentiate into Th1, Th2, Th17 or iTreg cells in the presence of specific cytokine environment [1].

Th1 cell differentiation requires IFN-γ and IL-12 that IFN-γ induces the Th1 key

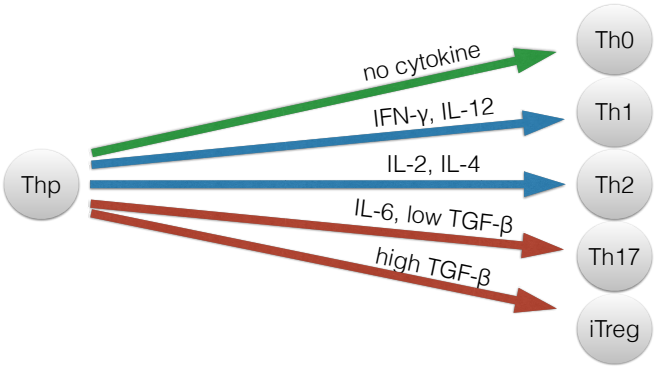

Figure 2.3: T helper (Th) cell lineages, mainly including Th1, Th2, Th17 and induced regulatory T (iTreg) cells. Th0 cells are the activated Th cells without differentiation.

TF, T-box 21 or TBX21 (T-bet), whereas IL-12 mediates the differentiation process [5]. The key TF of Th2 cells, trans-acting T-cell-specific transcription factor (GATA3), is up-regulated during Th2 cell differentiation in the presence of IL-2 and IL-4, but is down-regulated during Th1 cell differentiation [7]. Differentiation of Thp cells into Th17 cells requires IL-6 and low concentration of TGF-β to induce the key TF of Th17 cells, retinoic acid receptor-related orphan receptor gamma thymus (RORγt), which inhibits the key TF of iTreg cells, forkhead box P3 (FOXP3) [8, 9]. On the other hand, high concentration of TGF-β is required in iTreg cell differentiation in order to induce FOXP3 [8, 9]. From the required cytokine environment and regulation of TFs during Th cell differentiation, Th1-Th2 and Th17-iTreg are two pairs of reciprocally interconnected lineages of Th cells. Table 2.1 lists the lineages with their corresponding key TFs and required cytokine environment during the cell differentiation.

Further, signal transducer and activator of transcription (STAT), which is a family of TF, is also important in the signaling pathway that activation of STATs by cytokines is essential during Th cell differentiation [1]. The family member, STAT3, involved in Th17 cell differentiation is activated by IL-6, IL-21 and IL-23, in which STAT3 activation by IL-6 is responsible for induction of RORγt and FOXP3 down-regulation, and hence, this critical role determines the balance between Th17 and iTreg induction [1].

Lineages	Key TF	Signature Cytokines	Cytokine Environment
Th1	T-bet	IFN-γ	IFN-γ, IL-12
Th2	GATA3	IL-4, IL-5, IL-13	IL-2, IL-4
Th17	RORγt	IL-17A, IL-17F, IL-22	IL-6, low TGF-β
iTreg	FOXP3	IL-10, TGF-β	high TGF-β

Table 2.1: The T helper (Th) cell lineages and their corresponding key transcription factor (TF), signature cytokines and cytokine environment in the cell differentiation process.

2.3 RNA Sequencing

RNA sequencing (RNA-Seq), or Whole Transcriptome Shotgun Sequencing (WTSS), is a technology based on next-generation sequencing (NGS) to determine the RNA content of a cell or a cell population [13]. As other sequencing technologies, RNA-Seq can provide information about the complete snapshot of all the transcripts in a sample at a single time point. RNA-Seq can be used to quantify the number of reads originating from each of the mRNA molecules, and these read counts are typically modeled by the negative binomial distribution introduced in the later section [12, 13, 14, 15, 16]. Raw data contains reads of gene segments and has to be aligned. Reads are aligned to an annotated reference genome and those aligning to genes, including both exons and splice junctions, are counted. The process of mapping reads to a known genome and exons in known transcripts is called alignment [15].

Aligned read count data can be normalized to represent the abundance of genes. Reads per kilobase per million (RPKM) is one of the normalization method [22]. The number of reads C aligning to a gene is divided by the number of total mapped reads, called the library size (i.e. the total number of mapped reads) N, per one million. Then the resulting number is divided by the exon length (i.e. total length of all exons) L in base-pairs, per one thousand.

$$\text{RPKM} = \frac{10^9 \cdot C}{NL} \tag{2.1}$$

In our application, normalized RNA-Seq data provides the abundance of genes within a population, instead of a single cell. Therefore, we construct averaged population models with reaction rates, which are considered as averages over the population. Further, the number of cells are not in the consideration, and hence, cell division process can be neglected in the averaged population models.

Chapter 3

Mathematical Modeling of Biochemical Systems

In this chapter, we briefly discuss how ordinary differential equations (ODEs) can be used in the application of biochemical systems as the setting in which to represent the dynamical behavior by mathematical models. The presentation follows Chapter 1-3 and 7 of the book [23] with modified linear reaction examples. We also exemplify how ODE models can be constructed to describe the dynamics of biochemical reactions such as for gene regulatory networks (GRNs) in the studies [4, 5, 6, 7, 8, 9, 10, 11, 24]. Further, we introduce the general solution of the special form of ODEs, called matrix differential equations (MDEs) [25, 26, 27].

3.1 Ordinary Differential Equation (ODE) Modeling

The use of differential equations to model biochemical processes has a long history. One of the natural ways to model the dynamics of a biochemical pathway is using differential equations. Differential equations provide a highly expressive mathematical framework to model dynamic systems. It can describe reactions such as protein interactions in cell biology, population genetics in ecology and evolutionary biology and infectious disease transmission in mathematical epidemiology. In cell biology application reactions can be transcription, translation and degradation, etc [23]. The following gives a general approach to model molecule reactions in biological systems that can be applied to any application.

In this thesis, ordinary differential equations (ODEs) are used to describe the

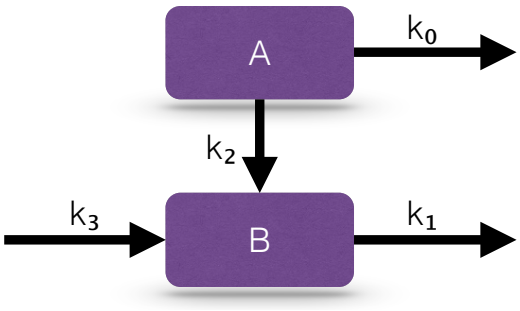

Figure 3.1: A simple reaction network involves degradation and production processes of A and B with linear reaction rates, k_0, k_1, k_2 and k_3.

dynamic behavior of reaction networks. Reaction rates are under the assumptions of spatial homogeneity and continuum hypothesis [23]. First, reactants are well-mixed, which means that all substances are equally distributed throughout the volume. Therefore, the rate of each reaction process is independent of space. Second, there are large numbers of reactants that the concentration of substances varies continuously according to the law of large numbers (LLN). Based on the above two assumptions, reaction rates can be described by the principle of mass action. The reaction rate is proportional to the product of the concentrations of reactants. In other words, the probability of a reaction occurring is proportional to the probability of reactants colliding with others. The kinetic order of a reactant is the exponent of reactant concentration. The constant of proportionality in the principle of mass action is called mass action rate constant, also called reaction rate.

In the following mathematical modeling, we let $[X]$ denote the concentration of reactant X and let $[X]_0$ denote the concentration at time $t = 0$, which is called the initial condition. Figure 3.1 shows an example of a simple reaction network, which is just an illustration consisting of degradation and production processes of molecules A and B, and it can be applied into any system. The following examples split the reaction network into two, the processes regulating A and regulating B.

Example 3.1.1. *Degradation Process of A*

$$A \xrightarrow{k_0}$$

This is a degradation process of A with the simple kinetic order = 1. Then the derivative of the function $[A]$ *with respect to time t, which is its rate of change,*

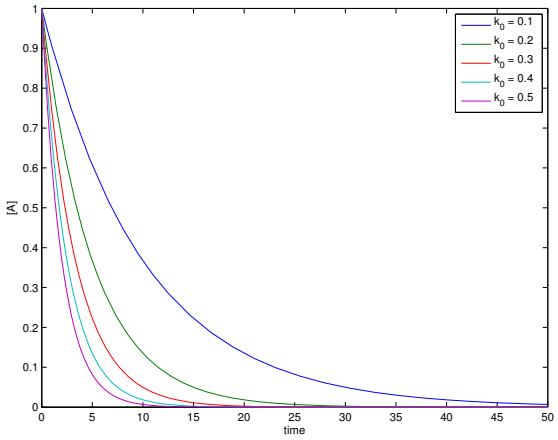

Figure 3.2: The exponential decreasing concentration $[A]$ in Equation (3.2) with different parameter $k_0 = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ given initial value $[A]_0 = 1$.

equals to the reaction rate $k_0[A]$*, as shown in Equation (3.1).*

$$\frac{d[A]}{dt} = -k_0[A] \tag{3.1}$$

$$[A] = [A]_0 e^{-k_0 t} \tag{3.2}$$

Equation (3.1) is an ordinary differential equation (ODE), with its analytical solution in Equation (3.2) describing the concentration over time t*. Figure 3.2 shows the dynamics of the solution with different parameter* k_0 *at* $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ *given initial value* $[A]_0 = 1$*.*

Example 3.1.2. *Multi Processes of* B

Let us now consider the full reaction network consisting of molecules A *and* B *in Figure 3.1. The simple reaction network involves three reactions of* B *including one degradation and two production processes, which can be split into three equations.*

$$\begin{aligned} B &\xrightarrow{k_1} \\ A &\xrightarrow{k_2} A + B \\ &\xrightarrow{k_3} B \end{aligned}$$

The first reaction describes B *decaying with the reaction rate* $k_1[B]$*. The second reaction is first-order production by* A *with the reaction rate* $k_2[A]$*, while the last reaction is zeroth-order production with the constant reaction rate* k_3*. The total rate of change of* B *is the sum of all the reaction rates. Therefore, the corresponding*

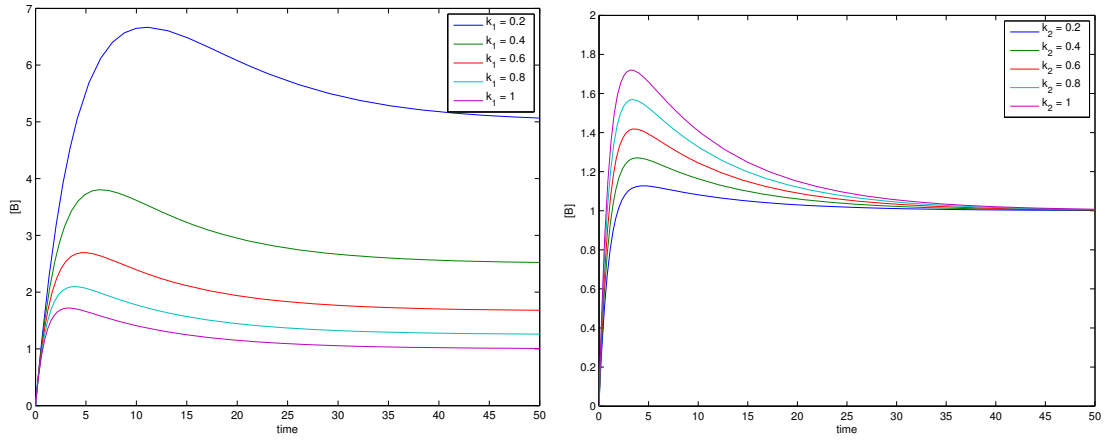

Figure 3.3: The left figure shows the concentration $[B]$ with different $k_1 = \{0.2, 0.4, 0.6, 0.8, 1\}$ given $k_0 = 0.1$, $k_2 = 1$, $k_3 = 1$, $[A]_0 = 1$ and $[B]_0 = 0$. The right figure shows the concentration $[B]$ with different $k_2 = \{0.2, 0.4, 0.6, 0.8, 1\}$ given $k_0 = 0.1$, $k_1 = 1$, $k_3 = 1$, $[A]_0 = 1$ and $[B]_0 = 0$.

ODE is

$$\frac{d[B]}{dt} = -k_1[B] + k_2[A] + k_3. \tag{3.3}$$

By substituting the analytical solution of $[A]$ *in Equation (3.2), the ODE changes from Equation (3.3) consisting of variables* $[A]$ *and* $[B]$ *into Equation (3.4) consisting of only one variable* $[B]$*, and then it can be solved analytically.*

$$\frac{d[B]}{dt} = -k_1[B] + k_2[A]_0 e^{-k_0 t} + k_3 \tag{3.4}$$

$$[B] = [B]_0 e^{-k_1 t} + e^{-k_1 t} \int_0^t e^{k_1 \tau}(k_2[A]_0 e^{-k_0 \tau} + k_3) d\tau \tag{3.5}$$

$$= \frac{k_2}{k_1 - k_0}[A]_0 e^{-k_0 t} + \Big([B]_0 - \frac{k_2}{k_1 - k_0}[A]_0 - \frac{k_3}{k_1}\Big) e^{-k_1 t} + \frac{k_3}{k_1} \tag{3.6}$$

Equations (3.1) and (3.3) are a coupled linear first-order non-homogeneous ODE system, with its analytical solution in Equations (3.2) and (3.6). Figure 3.3 shows the dynamics of the solution of $[B]$ *in Equation (3.6) with different parameters* k_1 *and* k_2 *at* $\{0.2, 0.4, 0.6, 0.8, 1\}$ *given initial values* $[A]_0 = 1$ *and* $[B]_0 = 0$ *as well as other parameters* $k_0 = 0.1$ *and* $k_3 = 1$*.*

3.2 Modeling Gene Regulatory Networks Using ODEs

In order to understand the dynamics of gene regulatory networks (GRNs), mathematical models have been used to describe molecular regulatory mechanisms [5, 6, 7, 8, 9, 10, 11]. In fact, the actual GRNs are unknown and complex consisting of many mechanisms describing genes regulate each other indirectly by their protein products. Because a huge number of molecules participates in gene regulatory mechanisms, not all reactants can be included in the models. Instead, GRN models can be constructed such as by describing transcription and translation events as well as degradation of some important transcription factors (TFs) in mRNA and protein levels to approximate the dynamics. The construction of ODE models for general GRNs is similar to the examples given in the previous section.

Example 3.2.1. *Simplest GRN*

Let us consider the simplest GRN shown in Figure 3.4 consisting of a single gene regulating its own activities. There are the basal transcription rate β and feedback transcription rate ψ regulated by protein. The translation of producing protein is regulated by mRNA with the rate λ. Further, both mRNA and protein have their own degradation rates δ_m and δ_p, respectively.

Let $[mRNA]$ and $[protein]$ denote the concentration of mRNA and protein levels of the gene, respectively. Then the corresponding ODEs are

$$\frac{d[mRNA]}{dt} = -\delta_m[mRNA] + \psi[protein] + \beta \tag{3.7}$$

$$\frac{d[protein]}{dt} = -\delta_p[protein] + \lambda[mRNA]. \tag{3.8}$$

Equations (3.7) and (3.8) are a coupled linear first-order non-homogeneous ODE system of the simplest GRN consisting of a single gene shown in Figure 3.4. The system can be solved analytically as the previous example. In fact, they are the special case of ODEs that can be written in matrix form called matrix differential equation (MDE) having a general solution.

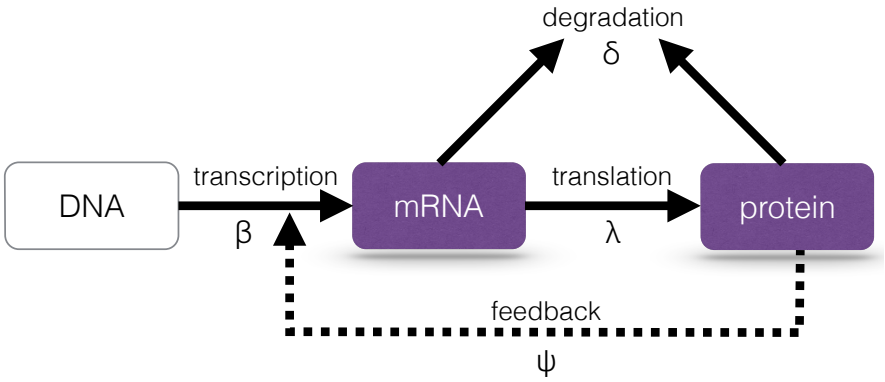

Figure 3.4: The simplest GRN consists of transcription and translation processes between mRNA and protein within a single gene as well as their degradation processes [24].

3.3 Solutions of Matrix Differential Equations

In general, an ordinary differential equation (ODE) can be written in the form

$$\frac{dx}{dt} = f(t, x), \tag{3.9}$$

where f is a given function of the two variables t and x, sometimes referred to as the rate function. We are interested in finding the differentiable function $x = x(t)$ that satisfies Equation (3.9) for all $t \geq 0$, called a solution. Unfortunately, there is no general method to solve Equation (3.9) analytically for an arbitrary function f. The analytic solution is available only for some special cases, for example those shown in Chapter 1-3 and 7 of the book [28]. Instead, we can approximate the solution by numerical methods such as Euler method, higher order Taylor series method, Heun's method, modified Euler's method and classical Runge-Kutta method that can be found in Chapter 8 of the book [28]. Note that some methods are more powerful than others depending on the order of the methods and the actual ODE structures. However, all these numerical methods have time-cumulative errors due to the calculation steps based on an initial value.

A matrix differential equation (MDE) is a special case of ODEs. The linear first-order ODEs with constant coefficients can be consider as MDE. The coefficients of ODEs are represented by the state companion matrix $\mathbf{A}$ and vector $\mathbf{b}$, whereas the solution $\mathbf{x} = \mathbf{x}(t)$ is a vector representing the multidimensional system response. In the following, we introduce both homogeneous and non-homogeneous MDEs with the general solutions. More importantly, the general solutions have no time-cumulative errors and the only time-dependent term is an important component

called the matrix exponential.

First, we consider a homogeneous MDE of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \tag{3.10}$$

where $\mathbf{A}$ is the $n \times n$ state companion matrix, $\mathbf{x}$ is a $n \times 1$ vector function and $\dot{\mathbf{x}}$ is the first derivative of $\mathbf{x}$ with respect to time t. Practically, if $\mathbf{A}$ has n distinct eigenvalues, $\lambda_1, \lambda_2, ..., \lambda_n$, then the general solution by linear combination is of the form

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2 + ... + c_n e^{\lambda_n t}\mathbf{v}_n, \tag{3.11}$$

where $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are the corresponding eigenvectors of $\mathbf{A}$ and $c_1, c_2, ..., c_n$ are constants, which depend on an initial value $\mathbf{x}(0)$. In principle, the solution can be also written in the matrix form

$$\mathbf{x}(t) = \mathbf{x}(0)e^{\mathbf{A}t}, \tag{3.12}$$

where $\mathbf{x}(0)$ is an initial value and $e^{\mathbf{A}\mathbf{t}}$ is the matrix exponential.

Further, we consider a non-homogeneous MDE of the form, which is added a $n \times 1$ constant parameter vector $\mathbf{b}$ on the right hand side of Equation (3.10)

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}. \tag{3.13}$$

In order to express the matrix solution, we transform a non-homogeneous MDE to a homogeneous MDE by considering the steady state or equilibrium $\mathbf{x}^*$ satisfying

$$\mathbf{A}\mathbf{x}^* + \mathbf{b} = 0. \tag{3.14}$$

Then the non-homogeneous MDE in Equation (3.13) becomes

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x} - \mathbf{x}^*), \tag{3.15}$$

where $\mathbf{x}^* = -\mathbf{A}^{-\mathbf{1}}\mathbf{b}$. Therefore, the matrix solution of Equation (3.13) is

$$\mathbf{x}(t) = \mathbf{x}^* + e^{\mathbf{A}t}(\mathbf{x}(0) - \mathbf{x}^*), \tag{3.16}$$

which is similar to Equation (3.12).

The matrix exponential $e^{\mathbf{A}}$ formally defined by the Taylor series in Equation (3.17) is the main part of the matrix solution in Equations (3.12) and (3.16).

$$e^{\mathbf{A}} = \sum_{l=0}^{\infty} \frac{\mathbf{A}^l}{l!} = 1 + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + ... \tag{3.17}$$

There are several methods to calculate or approximate matrix exponentials such as Taylor series expansion [25], matrix decomposition method [25], Putzer algorithm [26] and the scaling and squaring method [27].

Chapter 4

Bayesian Analysis

In this chapter, we present the idea of Bayesian analysis following Chapter 1-3 of the book [17] and also statistical models including likelihood functions and prior distributions. These statistical models are important in Bayesian analysis, which provides a way to draw conclusions from observation based on the assumption of the distributions. Further, since different dynamical models are often generated to describe the same system, Bayesian inference can provide a systemic approach to select alternative models and rank them with a mathematical probability called the model posterior or the ratio called Bayes factor between model evidences [29]. Moreover, we present the concept of posterior predictive distribution that we use in the context of ODE models to predict dynamic behavior [30].

4.1 Concepts on Bayesian Inference

Bayesian inference is an important technique in statistics to draw conclusions from observed data. In statistical content, hypotheses can be expressed through probability distributions and the probability distributions depend on parameters. The analysis is a procedure to estimate parameters of an underlying distribution based on the observation. The method is mathematically based on Bayes' rule to to update the probability for a hypothesis as evidence is acquired.

The following three bullet points present the steps to the process of Bayesian data analysis which is formulated in the book [17].

- *"Setting up a full probability model - a joint probability distribution for all observable and unobservable quantities in a problem. The model should be*

consistent with knowledge about the underlying scientific problem and the data collection process."

- *"Conditioning on observed data: calculating and interpreting the appropriate posterior distribution - the conditional probability distribution of the unobserved quantities of ultimate interest, given the observed data."*
- *"Evaluating the fit of the model and the implications of the resulting posterior distribution: how well does the model fit the data, are the substantive conclusions reasonable, and how sensitive are the results to the modeling assumptions in step 1? In response, one can alter or expand the model and repeat the three steps."*

Bayes' rule is the key part of Bayesian inference. It states mathematically as the equation about two events A and B in the following

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \tag{4.1}$$

$$= \frac{p(B|A)p(A)}{p(B)}, \tag{4.2}$$

where $p(A|B)$ is a conditional probability of A given that B is true, $p(B|A)$ is the reverse conditional probability, $p(A \cap B)$ is joint probability of both events, and lastly $p(A)$ and $p(B)$ are the probabilities of A and B respectively. This is a general formula and the following represents it on Bayesian.

In Bayesian data analysis, we simply replace A and B by parameter θ and data set D, respectively. Then Bayes' rule becomes

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{4.3}$$

where the posterior distribution $p(\theta|D)$ is a product of the likelihood $p(D|\theta)$ and prior $p(\theta)$ divided by the marginal likelihood $p(D)$.

The posterior distribution $p(\theta|D)$ is the probability distribution of the unknown quantity conditional on the observations. After data is observed, parameter θ of unknown quantity does not only depend on the current knowledge, which is presented by the prior distribution $p(\theta)$, but also based on the information represented by the likelihood from the data set D.

The marginal likelihood $p(D)$, or integrated likelihood, is a likelihood function in which parameter has been marginalized, and therefore, it is a normalizing constant that gives evidence. Since the parameter θ here is considered as a continuous

parameter, the marginal likelihood $p(D)$ is the integral of parameter θ, i.e. the expectation of the likelihood with respect to the prior, as

$$p(D) = \int_{\theta} p(D|\theta)p(\theta)d\theta. \tag{4.4}$$

4.1.1 Likelihood Functions

In modern statistics, likelihood is one of the most fundamental concepts [17]. A likelihood function $\mathcal{L}(\theta|y)$ is a function, which equals to a density $p(y|\theta)$ for a conditional probability of y given θ, of a parameter θ for fixed data y. If we consider that the likelihood function is a discrete probability distribution and let Y be a random variable with the discrete probability distribution p depending on a parameter θ, then

$$\mathcal{L}(\theta|y) = p(y|\theta) = P(Y = y|\theta). \tag{4.5}$$

In other words, likelihood is the hypothetical probability that an event which has already occurred would yield a specific outcome. It allows us to estimate unknown parameters based on known outcomes. Likelihood is a tool for summarizing the evidence of data about unknown parameters.

In many applications, independent observations are collected and therefore the likelihood function is a product over individuals. We let N be the number of observations and $\mathbf{y} \in \{y_1, ..., y_N\}$ be the set of observations, then

$$p(\mathbf{y}|\theta) = \prod_{n=1}^{N} \mathcal{L}(\theta|y_n). \tag{4.6}$$

We assume that observed data y follows a specific parametric distribution which is the likelihood with parameter θ. In other words, the likelihood determines the statistical model and observations are assumed to be drawn from this distribution. For example, it is very common to assume a continuous probability distribution as a normal distribution, which can be found in the appendix. On the other hand, Poisson distributions, exponential distributions, gamma distributions, binomial distributions and negative binomial distributions [12, 13, 14, 15, 16] are often used in discrete probability distribution as an assumption.

4.1.2 Prior Distributions

A prior probability distribution, or simply called prior, is another fundamental concept in Bayesian statistics [17]. The prior $\rho(\theta)$ is a probability distribution of

parameter θ before data y is taken into account. It is an assumption of uncertainties based on current knowledge that expresses beliefs without the consideration of the data.

In this thesis, the parameters are the coefficients in the systems representing reaction rates and differentiation population proportion constants. In our mode, there is no prior dependencies between parameters, and hence, prior collects the information of parameters as a product that is similar to multi-observations of the likelihood function. We let d be the number of parameters and $\boldsymbol{\theta} = \{\theta_1, ..., \theta_d\}$ be the set of parameters, then

$$p(\boldsymbol{\theta}) = \prod_{m=1}^{d} \rho(\theta_m). \tag{4.7}$$

There are two types of priors, informative and uninformative [17]. The former gives a stronger assumption of the uncertainties especially when some information is available to restrict domination by the likelihood on a posterior distribution, the latter is a relatively flat distribution with a weaker assumption to minimize the impact on the posterior distribution. In this thesis, we use a proper uninformative prior that is used to constrain parameter values in a physically meaningful range. In general, standard normal distributions, standard log-normal distributions and uniform distributions are often good choices as uninformative priors. The log-normal distributions and uniform distributions can be found in the appendix.

4.2 Model Ranking

In many applications, there are different alternative models to describe data. First, we consider Equations (4.3) and (4.4) for a given model M_k, then they become the parameter posterior $p(\theta|D, M_k)$ and marginal likelihood $p(D|M_k)$

$$p(\theta|D, M_k) = \frac{p(D|\theta, M_k)p(\theta|M_k)}{p(D|M_k)} \tag{4.8}$$

$$p(D|M_k) = \int_{\theta} p(D|\theta, M_k)p(\theta|M_k)d\theta, \tag{4.9}$$

where $p(D|\theta, M_k)$ and $p(\theta|M_k)$ are now called the data likelihood and parameter prior, respectively.

Second, we consider a K-model class $\mathbf{M} = \{M_1, M_2, ..., M_K\}$ with the model prior distribution $p(M_k)$, then the model posterior $p(M_k|D)$ is obtained by Bayes'

rule

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{\sum_{i=1}^{K} p(D|M_i)p(M_i)}, \tag{4.10}$$

where $k = 1, 2, ...K$. Be aware that, besides parameter prior $p(\theta|M_k)$ the model needs to specify the prior for the alternative models and this have a joint distribution called model prior $p(M_k)$. Consequently, the posterior distribution is called model posterior $p(M_k|D)$ which is used for model ranking and prediction shown in the following.

The marginal likelihood, which represents the model evidence given observed data, is an important component to measure evidential preferences to compare alternative models and to select the most suitable model [18, 19, 20, 29, 30]. In fact, model posteriors and Bayes factor are common tools to compare models using the resulting marginal likelihoods. In practice, it is often to assume uniform model prior, i.e. $p(M_k) = 1/K, \forall k = 1, ..., K$. Then Equation (4.10) becomes

$$p(M_k|D) = \frac{p(D|M_k)}{\sum_{i=1}^{K} p(D|M_i)}. \tag{4.11}$$

Therefore, marginal likelihoods can directly show the evidence of models that the higher marginal likelihood means the higher model posterior, and the highest model posterior $p(M_k|D)$ illustrates the model M_k is the best model to describe the data D.

On the other hand, two alternative models can be also compared by Bayes factor $B_{kk'}$, which is simply the ratio of the two respective posteriors for models M_k and $M_{k'}$ in Equation (4.12).

$$B_{kk'} = \frac{p(M_k|D)}{p(M_{k'}|D)} \tag{4.12}$$

$$= \frac{p(D|M_k)p(M_k)}{p(D|M_{k'})p(M_{k'})} \tag{4.13}$$

$$= \frac{p(D|M_k)}{p(D|M_{k'})} \tag{4.14}$$

Equations (4.13) and (4.14) show that Bayes factor is also the ratio of the two respective marginal likelihoods by assuming uniform model prior.

The interpretation of the Bayes factor as evidence support categories is shown in Table 4.1 [29, 31]. Clearly, $B_{kk'} > 1$ favors model M_k than $M_{k'}$ simply because $p(D|M_k) > p(D|M_{k'})$, but it is not worth more than a bare mention while $1 < B_{kk'} < 3.2$ due to the similarity between two models. Most importantly, $3.2 < B_{kk'} < 10$ substantially favors model M_k and $10 < B_{kk'} < 100$ strongly favors model M_k, while $B_{kk'} > 100$ decisively favors model M_k.

$B_{kk'}$	$\log B_{kk'}$	$\log_{10} B_{kk'}$	Evidence support
$(1, 3.2)$	$(0, 1.2)$	$(0, 0.5)$	Not worth more than a bare mention
$(3.2, 10)$	$(1.2, 2.3)$	$(0.5,\ 1)$	Substantial
$(10, 100)$	$(2.3, 4.6)$	$(1, 2)$	Strong
> 100	> 4.6	> 2	Decisive

Table 4.1: Interpretation of the Bayes factor as evidence support categories according to the studies [29, 31].

4.3 Posterior Predictive Distributions

This section follows the ideas and notations of the article [30] to present how Bayesian inference can be used in posterior predictions. Consider a one-dimensional N-point observed data set $D = \{\mathbf{y}, \mathbf{t}\}$ such that $\mathbf{y} \in \mathbb{R}^N$ and $\mathbf{t} \in \mathbb{R}^N$, where $\mathbf{y}$ and $\mathbf{t}$ are the observables and observed time points, respectively. The prediction of unobserved variables $\mathbf{y}^*$ at time t^* is of a distribution called the averaged prediction distribution $p(\mathbf{y}^*|t^*, D)$, which is obtained by averaging the predictive distribution $p(\mathbf{y}^*|t^*, D, M_k)$ for each model with respect to the discrete model posterior $p(M_k|D)$.

$$p(\mathbf{y}^*|t^*, D) = \sum_{k=1}^{K} p(\mathbf{y}^*|t^*, D, M_k)p(M_k|D) \tag{4.15}$$

Before averaging, the predictive posterior $p(\mathbf{y}^*|t^*, D, M_k)$ for each model is the integral of the model specific predictive likelihood $p(\mathbf{y}^*|t^*, \theta, M_k)$ with respect to the parameter posterior $p(\theta|D, M_k)$.

$$p(\mathbf{y}^*|t^*, D, M_k) = \int_{\theta} p(\mathbf{y}^*|t^*, \theta, M_k)p(\theta|D, M_k)d\theta \tag{4.16}$$

Note that if there is a model, say M_k, dominating the posterior, i.e. $p(M_k|D) \approx 1$, then Equation (4.15) becomes

$$p(\mathbf{y}^*|t^*, D) = p(\mathbf{y}^*|t^*, D, M_k), \tag{4.17}$$

where the averaged prediction distribution $p(\mathbf{y}^*|t^*, D)$ is simply the predictive posterior $p(\mathbf{y}^*|t^*, D, M_k)$ of the model M_k.

Chapter 5

Bayesian Computation

In this chapter, we outline the computation steps of Markov chain Monte Carlo (MCMC) sampling algorithms that are often used to carry out the posterior analysis. MCMC plays an important role in parameter posterior sampling because it requires a large scale sampling in high dimensional space, which is impossible to calculate by hands but computers instead. We provide algorithms, Metropolis-Hastings (MH) [32, 33] and population-based MCMC (popMCMC) [18, 19, 20, 34, 35], in parameter estimation to draw samples from target distributions. Further, marginal likelihood estimation uses those samples to rank alternative models in order to select the most suitable model describing observed data. The power posterior estimator [18, 19, 20] is a marginal likelihood estimator using samples drawn from popMCMC.

5.1 Markov Chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) is a method consisting of two concepts, Markov chain and Monte Carlo [36], to approximate target distributions which are the parameter posterior distributions in our application. Ideally, it generates random draws such that Markov chain converges to target distribution as its equilibrium. This method is able to draw samples from a high dimensional distribution to do the approximation, while sampling directly from target distribution is impossible.

Monte Carlo is a method to solve problems by generating random numbers and observing the fraction of obeying some properties. There are three main problem classes of Monte Carlo simulation and they are optimization, numerical integra-

tion and sampling from a probability distribution. For example, the importance sampling and rejection sampling are two common methods of Monte Carlo integration [36]. In the following, we present algorithms to carry out MCMC sampling. MCMC constructs a Markov chain that converges to the equilibrium distribution, also called stationary distribution. Markov chain is a random process that transitions from one to another depends on only the current state, which is characterized as memoryless. At the equilibrium, the process is reversible and satisfies the detailed balance equation from the current state θ to the next state θ^*

$$q(\theta^*|\theta)r(\theta) = q(\theta|\theta^*)r(\theta^*), \tag{5.1}$$

where $q(\theta^*|\theta)$ is Markov transition probability called the proposal distribution from θ to θ^* and $r(\theta)$ is the target distribution that the Markov chain should converge to. The detailed introduction of MCMC can be found in Chapter 1 of the book [36].

5.2 MCMC Sampling Algorithms

Many MCMC algorithms have been developed during recent years. The common methods are Metropolis-Hastings (MH), adaptive Metropolis (AM), Gibbs sampling, multiple-try Metropolis and reversible-jump MCMC algorithms. Different sampling methods have advantages on different applications and some of them are more advanced than others. Many other algorithms have been invented in order to reduce the convergence time for efficiency. In this thesis, besides the basic method MH algorithm [32, 33], population-based MCMC (popMCMC) [18, 19, 20, 34, 35] based on MH algorithm are introduced.

5.2.1 Metropolis-Hastings Algorithm

Metropolis-Hastings (MH) algorithm shown in Algorithm 1 is the simplest way of the family of the MCMC sampling algorithms [32, 33]. Remind that MCMC uses Monte Carlo method to draw random samples to observe the fraction of obeying some properties. The algorithm is simply a random walk with an acceptance rule to decide that the sample is accepted and move, or rejected and stay.

First, an initial sample $\theta^{(1)}$ is drawn randomly, which is chosen to be drawn from the prior $p(\theta)$ in our algorithm. A random walk starts from the initial sample $\theta^{(1)}$ and then Markov chain samples $\theta^{(2)}, ..., \theta^{(J)}$ are drawn one-by-one from the proposal distribution $q(\theta^*|\theta)$, where θ and θ^* are the current state and proposed

draw $\theta^{(1)} \sim p(\theta)$
for $j = 2$ **to** $j = J$ **do**
  draw $\theta^* \sim q(\theta^*|\theta^{(j-1)})$
  draw $u \sim U[0, 1]$
  if $u < A = \frac{q(\theta^{(j-1)}|\theta^*)}{q(\theta^*|\theta^{(j-1)})} \frac{r(\theta^*)}{r(\theta^{(j-1)})}$ **then**
    accept: $\theta^{(j)} = \theta^*$
  else
    reject: $\theta^{(j)} = \theta^{(j-1)}$
  end
end

Algorithm 1: Metropolis-Hastings (MH) algorithm

state respectively. According to the acceptance rule, samples are accepted by the acceptance probability

$$A = \min\left\{1, \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \frac{r(\theta^*)}{r(\theta)}\right\}, \tag{5.2}$$

where $r(\theta)$ is the target distribution.

In the computation shown in Algorithm 1, a uniformly distributed random variable u is generated to choose the acceptance. Given θ and θ^*, if $u < A$, then the proposed state θ^* is accepted and the Markov chain moves to the proposed sample. Otherwise, the chain stays at the current sample θ. A sequence of the samples is then obtained by repeating J iterations following the steps above.

The MH algorithm is simple and mainly controlled by the proposal distribution $q(\theta^*|\theta)$. For many applications, a symmetric random walk proposal such that $q(\theta^*|\theta) = q(\theta|\theta^*)$, is chosen for simplicity. Then the acceptance probability A no longer depends on the proposal, but only the target distribution. Therefore, Equation (5.2) becomes

$$A = \min\left\{1, \frac{r(\theta^*)}{r(\theta)}\right\}. \tag{5.3}$$

This simplified version of MH algorithm is called Metropolis algorithm. A symmetric proposal such as normal distribution is often chosen. In the following algorithms, the proposal distributions are assumed to be normally distributed.

In Bayesian data analysis, the posterior distribution is the target distribution.

$$r(\theta) = p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{5.4}$$

In fact, the actual posterior distribution is not needed and difficult to be obtained because of the normalizing constant, the marginal likelihood $p(D)$. During MCMC sampling, Equation (5.5) shows that only the unnormalized posteriors, which are the products of the likelihood and prior, are needed to calculate the acceptance probability in Equation (5.3).

$$\frac{r(\theta^*)}{r(\theta)} = \frac{p(D|\theta^*)p(\theta^*)}{p(D|\theta)p(\theta)} \tag{5.5}$$

5.2.2 Population-based MCMC Algorithm

Population-based MCMC (popMCMC) algorithm shown in Algorithm 2 is an advanced method based on MH algorithm and thermodynamics [18, 19, 20, 34, 35]. This is a more powerful method to use power posterior distributions in Bayesian inference defined in Equation (5.6) to approximate target distributions. Instead of drawing samples from the posterior distribution only, popMCMC simultaneously draws samples from several distributions, the power posterior distributions. The power posterior distributions $p_\tau(\theta_\tau|D)$ with temperatures τ defined from 0 to 1 are constructed bridging from prior to posterior.

$$p_\tau(\theta_\tau|D) = \frac{p(D|\theta_\tau)^\tau p(\theta_\tau)}{\int_{\theta_\tau} p(D|\theta_\tau)^\tau p(\theta_\tau) d\theta_\tau} \propto p(D|\theta_\tau)^\tau p(\theta_\tau) \tag{5.6}$$

Clearly, this is the prior distribution $p(\theta)$ if $\tau = 0$, while this is the posterior distribution $p(\theta|D)$ if $\tau = 1$.

In popMCMC, the temperature $\tau \in [0, 1]$ is discretized in to τ_n, where $n = 1, 2, ...$ There are several distributions whose number equals to the number of temperatures. To begin the sampling algorithm, random walk starts from initial samples drawn at all temperatures. During the sampling, each iteration moves only one step along one of the temperature. It is either a local move or global move that the local move to walk along each temperature or global move to swap samples between temperatures are randomly chosen. If the uniformly distributed random variable $v < 0.5$, the iteration is chosen to be a local move, otherwise, a global move. Then a temperature τ_n is randomly chosen for the sample draw if the move is local, while a pair of neighbor temperatures $\{\tau_n, \tau_{n+1}\}$ is randomly chosen if this is a global move.

During local moves, samples $\theta^*_{\tau_n}$ are generated from proposal distributions $q_{\tau_n}(\theta^*_{\tau_n}|\theta_{\tau_n})$ and accepted according to the acceptance rule with acceptance probability A_{τ_n} defined in Equation (5.7) at each temperature τ_n. These usual walks simply follow

draw $\theta_{\tau_n}^{(1)} \sim p(\theta)$ at each temperature τ_n
for $j = 2$ **to** $j = J$ **do**
  draw $v \sim U[0,1]$
  if $v < 0.5$ **then**
    randomly choose one temperature τ_n
    draw $\theta_{\tau_n}^{*} \sim q_{\tau_n}(\theta_{\tau_n}^{*}|\theta_{\tau_n}^{(j-1)})$
    draw $u \sim U[0,1]$
    if $u < A_{\tau_n} = \frac{p(D|\theta_{\tau_n}^{*})^{\tau_n} p(\theta_{\tau_n}^{*})}{p(D|\theta_{\tau_n})^{\tau_n} p(\theta_{\tau_n})}$ **then**
      accept: $\theta_{\tau_n}^{(j)} = \theta_{\tau_n}^{*}$
    else
      reject: $\theta_{\tau_n}^{(j)} = \theta_{\tau_n}^{(j-1)}$
    end
  else
    randomly choose a pair of neighbor temperatures $\{\tau_n, \tau_{n+1}\}$
    draw $u \sim U[0,1]$
    if $u < A'_{\tau_n} = \frac{p(D|\theta_{\tau_{n+1}})^{\tau_n}}{p(D|\theta_{\tau_n})^{\tau_n}} \frac{p(D|\theta_{\tau_n})^{\tau_{n+1}}}{p(D|\theta_{\tau_{n+1}})^{\tau_{n+1}}}$ **then**
      accept: $\theta_{\tau_n}^{(j)} = \theta_{\tau_{n+1}}^{(j-1)}$ and $\theta_{\tau_{n+1}}^{(j)} = \theta_{\tau_n}^{(j-1)}$
    else
      reject: $\theta_{\tau_n}^{(j)} = \theta_{\tau_n}^{(j-1)}$ and $\theta_{\tau_{n+1}}^{(j)} = \theta_{\tau_{n+1}}^{(j-1)}$
    end
  end
end

Algorithm 2: Population-based MCMC (popMCMC) algorithm

the steps of MH algorithm.

$$A_{\tau_n} = \frac{p_{\tau_n}(\theta_{\tau_n}^{*}|D)}{p_{\tau_n}(\theta_{\tau_n}|D)} = \frac{p(D|\theta_{\tau_n}^{*})^{\tau_n} p(\theta_{\tau_n}^{*})}{p(D|\theta_{\tau_n})^{\tau_n} p(\theta_{\tau_n})} \tag{5.7}$$

On the other hand, there are samples during global moves swapping between neighbor temperatures $\{\tau_n, \tau_{n+1}\}$ according to the acceptance rule with acceptance probability A'_{τ_n} defined in Equation (5.8). The global moves help linking all power posterior distributions together due to the similarity between each pair of neighbor distributions.

$$A'_{\tau_n} = \frac{p_{\tau_n}(\theta_{\tau_{n+1}}|D)}{p_{\tau_n}(\theta_{\tau_n}|D)} \frac{p_{\tau_{n+1}}(\theta_{\tau_n}|D)}{p_{\tau_n}(\theta_{\tau_{n+1}}|D)} = \frac{p(D|\theta_{\tau_{n+1}})^{\tau_n}}{p(D|\theta_{\tau_n})^{\tau_n}} \frac{p(D|\theta_{\tau_n})^{\tau_{n+1}}}{p(D|\theta_{\tau_{n+1}})^{\tau_{n+1}}} \tag{5.8}$$

When the number of temperatures increases, a higher number of distributions are estimated, but at the same time, the global acceptance rate is higher due to the

higher similarity between the neighbor distributions. Since there are more than one distribution estimated, the convergence time is much longer. For computational efficiency, there are mainly two ways to accelerate the sampling, where one focuses on the local move and another focuses on the global move.

To speed up popMCMC sampling from the local moves, one can adopt the idea from AM algorithm [37], using the adaptation in the proposal distributions. In practice, it is difficult to choose good proposal distributions and fixed proposals are not efficient to search high probability regions at an early stage. The proposal distribution at each temperature τ_n can be updated by recalculating the covariance matrix

$$\Sigma_{\tau_n} = s_d cov(\theta_{\tau_n}^{(j-J_s)}, ..., \theta_{\tau_n}^{(j-1)}) + s_d \epsilon I_d, \tag{5.9}$$

where $s_d = \frac{2.38^2}{d}$ and d is the dimension of parameter vector θ. However, this is different with AM algorithm because AM does not construct a Markov chain but has the correct ergodic properties [37]. Adding adaptation to popMCMC may cause ergodic properties changed. Therefore, adaptation is only applied during the burn-in period and the covariance matrix is fixed after warm-up, in order to keep the Markov chain properties of those saved samples. This substantiates the saved samples form a Markov chain that satisfies the detailed balance equation.

On the other hand, the choice of temperature schedule $\tau_1, \tau_2, ...$ is a key point to accelerate from the global moves. Remind that the strength of popMCMC is adding the global moves swapping samples between neighbor temperatures. If two neighbor distributions are similar, there is a high probability to accept the swapping, otherwise, lower acceptance global moves have no advantages in sampling. The temperature schedule depends on the models. Temperature schedules with decreasing intervals, i.e. differences between higher temperatures is smaller than those lower temperatures, are often chosen because of the unknown posterior distribution at $\tau = 1$. This refers to the distributions with higher temperatures are more difficult and complicated to draw samples from, while those distributions with lower temperatures are easier due to the known prior at $\tau = 0$.

5.3 Power Posterior Estimator

The estimation of the marginal likelihood function defined in Equation (4.9) is often a difficult task because of its integration of nonlinear functions in high dimensional space. There are many existing marginal likelihood estimators such as the prior arithmetic mean (PAM) and posterior harmonic mean (PHM) estimators

[29]. In this thesis, we use the power posterior estimator [18, 19, 20] to estimate marginal likelihoods of all alternative models.

The power posterior estimator is an advanced method to estimate marginal likelihoods based on thermodynamic integration [18, 19, 20]. It uses samples from the power posterior distributions with temperature τ from 0 to 1 that bridge from the prior to posterior distribution. The samples can be obtained by drawing from individual distributions, which means several simulations for each temperature, or directly using population-based MCMC (popMCMC) algorithm. This approach is simple to implement by the following calculation. First, the power posteriors defined in Equation (5.6) for a given model M_k are

$$p_\tau(\theta_\tau|D, M_k) = \frac{p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k)}{\int_{\theta_\tau} p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k) d\theta_\tau} \propto p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k), \quad (5.10)$$

where $\int_{\theta_\tau} p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k) d\theta_\tau$ is an normalizing constant, denoted by $z(D|\tau, M_k)$ as a function of temperature τ. We then take the derivative of $\log z(D|\tau, M_k)$ with respect to τ,

$$\begin{aligned}
\frac{d}{d\tau} \log z(D|\tau, M_k) &= \frac{1}{z(D|\tau, M_k)} \frac{d}{d\tau} z(D|\tau, M_k) && (5.11)\\
&= \frac{1}{z(D|\tau, M_k)} \frac{d}{d\tau} \int_{\theta_\tau} p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k) d\theta_\tau && (5.12)\\
&= \frac{1}{z(D|\tau, M_k)} \int_{\theta_\tau} p(D|\theta_\tau, M_k)^\tau \log\{p(D|\theta_\tau, M_k)\} p(\theta_\tau|M_k) d\theta_\tau && (5.13)\\
&= \int_{\theta_\tau} \frac{p(D|\theta_\tau, M_k)^\tau p(\theta_\tau|M_k)}{z(D|\tau, M_k)} \log\{p(D|\theta_\tau, M_k)\} d\theta_\tau && (5.14)\\
&= \mathbb{E}_{\theta_\tau|D,\tau,M_k}[\log\{p(D|\theta_\tau, M_k)\}], && (5.15)
\end{aligned}$$

which is the expectation of the data likelihood in log-scale taken with respect to the parameter θ_τ. Then the natural logarithm of the marginal likelihood is

$$\begin{aligned}
\log p(D|M_k) &= \log \frac{z(D|\tau = 1, M_k)}{z(D|\tau = 0, M_k)} && (5.16)\\
&= \int_0^1 \mathbb{E}_{\theta_\tau|D,\tau,M_k}[\log\{p(D|\theta_\tau, M_k)\}] d\tau. && (5.17)
\end{aligned}$$

The integral in Equation (5.17) can be approximated by the trapezoidal rule numerically. The temperature domain $\tau \in [0, 1]$ is divided into $N_\tau - 1$ intervals according to the temperature schedule τ_n for $n = 1, ..., N_\tau$. Therefore,

$$\int_0^1 f(\tau) d\tau \approx \frac{1}{2} \sum_{n=1}^{N_\tau - 1} (\tau_{n+1} - \tau_n)[f(\tau_{n+1}) + f(\tau_n)], \quad (5.18)$$

where

$$f(\tau) = \mathbb{E}_{\theta_\tau|D,\tau,M_k}[\log\{p(D|\theta_\tau, M_k)\}] \tag{5.19}$$

$$\approx \frac{1}{J}\sum_{j=1}^{J} \log\{p(D|\theta_\tau^{(j)}, M_k)\}, \tag{5.20}$$

where J is the number of samples $\theta_\tau^{(j)}$ drawn from the power posterior $p_\tau(\theta_\tau|D, M_k)$.

5.4 Monitoring MCMC Convergence

In practice, there are a few steps after MCMC sampling to monitor convergence of the obtained samples. MCMC sampling requires a warm-up process called burn-in, which is the period before Markov chain converging to the equilibrium distribution. Those samples during the burn-in period are discarded and the saved samples after burn-in period are considered as the samples drawn from the equilibrium distribution. Therefore, the sample number has to be large enough to make sure convergence and the book [17] suggests that half of the samples, i.e. $J/2$ iterations, are discarded as the warm-up.

In statistics, independent and identically distributed (i.i.d.) random variables mean each has the same probability distribution as all of them are mutually independent. Samples generated from Markov chains can be strongly autocorrelated, which means they are not i.i.d.. In order to approximate a distribution by i.i.d. samples, we need another process called thinning to reduce autocorrelation by discarding all but saving every k^{th} iteration, where k is chosen according to the autocorrelation of the sequence.

Autocorrelation, also called serial correlation, is a mathematical tool to describe the similarity of observations by their reporting patterns. Statistically, the autocorrelation of a random process represents the correlation between two time points or the time lag. For stationary process, i.e. mean and variance are time-independent, the autocorrelation depends only on the time lag. In MATLAB, there is a built-in function `autocorr` estimating the autocorrelation of a stochastic sequence. Consider a T-step stochastic process y_t, the estimated autocorrelation for lag s, which is the measurement between y_t and y_{t+s}, is

$$r_s = \frac{c_s}{c_0}, \tag{5.21}$$

where $c_s = \frac{1}{T-1}\sum_{t=1}^{T-s}(y_t - \bar{y})(y_{t+s} - \bar{y})$ and c_0 is the sample variance of the time series.

On the other hand, sampling usually needs other techniques such as generating independent chains to confirm a reliable equilibrium distribution. This is because the approximation result can be wrong that a chain converges to a distribution locally instead of discovering the whole target distribution globally. For many applications of unknown target distributions, one way to check the result is generating independent Markov chains to determine the convergence range. Several chains converge to the same equilibrium distribution can be a reliable evidence to verify the results.

After that, we can use potential scale reduction introduced in Chapter 11 of the book [17] to check convergence of samples by mixing and stationarity. This is a simple approach by splitting chains in half and checking the resulting sequences. Suppose there are K independent chains, each of J iterations, and half of the sequences are discarded as the burn-in. By splitting the remaining sequences into half, there are $\tilde{K} = 2K$ chains, each of $\tilde{J} = J/4$ iteration. Let θ_{jk} denote the estimand for iteration $j = 1, 2, ..., \tilde{J}$ and chain $k = 1, 2, ..., \tilde{K}$. Then the between-sequence variance B and within-sequence variance W are

$$B = \frac{\tilde{J}}{\tilde{K} - 1} \sum_{\tilde{k}=1}^{\tilde{K}} (\bar{\theta}_{\cdot\tilde{k}} - \bar{\theta}_{\cdot\cdot})^2 \tag{5.22}$$

$$W = \frac{1}{\tilde{K}} \sum_{\tilde{k}=1}^{\tilde{K}} s_{\tilde{k}}^2, \tag{5.23}$$

where the within-sequence mean $\bar{\theta}_{\cdot\tilde{k}} = \frac{1}{\tilde{J}} \sum_{\tilde{j}=1}^{\tilde{J}} \theta_{\tilde{j}\tilde{k}}$, the overall mean $\bar{\theta}_{\cdot\cdot} = \frac{1}{\tilde{K}} \sum_{\tilde{k}=1}^{\tilde{K}} \bar{\theta}_{\cdot\tilde{k}}$ and the variance of each sequence $s_{\tilde{k}}^2 = \frac{1}{\tilde{J}-1} \sum_{\tilde{j}=1}^{\tilde{J}} (\theta_{\tilde{j}\tilde{k}} - \bar{\theta}_{\cdot\tilde{k}})^2$. The potential scale reduction is estimated by

$$\hat{R} = \sqrt{\frac{\hat{var}^+(\theta|D)}{W}}, \tag{5.24}$$

where the marginal posterior variance of the estimand $\hat{var}^+(\theta|D) = \frac{\tilde{J}-1}{\tilde{J}} W + \frac{1}{\tilde{J}} B$. In principle, $\hat{R} \to 1$ as $\tilde{J} \to \infty$. Hence, if the potential scale reduction declines to near 1, then the samples are well mixed and converge to the equilibrium distribution.

Chapter 6

Approximative Modeling Approaches for T Helper 17 Cell Differentiation

In this chapter, we present the models of the study for the T helper (Th) 17 cell differentiation. Th0 and Th17 cell networks determine the cell type specific dynamics of Th0 and Th17 cells respectively that both of them have similar but different dynamics because Th17 cell network are constructed based Th0 cell network. Alternative models, including one homogeneous population model (M_1) and two heterogeneous population models (M_2, M_3), are constructed based on the networks of Th0 and Th17 cells. Further, we provide efficient evaluation of explicit solutions for matrix differential equations (MDEs) and statistical models of RNA sequencing (RNA-Seq) data as well as computational implementation.

6.1 Motivation of Modeling Heterogeneity with Cell Type Specific Dynamics

Homogeneous cell population is usually assumed in cell differentiation modeling approaches. In reality, cell differentiation efficiency to any of the lineages is not perfect, and therefore, a cell population is heterogeneous instead of homogeneous. The heterogeneous cell population structured models consist of different cell types having their own reaction networks and kinetics. We assume gene regulatory networks (GRNs) of cell types are independent of each others, which means there is no cell interaction between subtypes, and then each ordinary differential equation

(ODE) system describes the reactions of a specific cell type independently. Since all subtypes are differentiated from precursor cells, reaction networks are different but similar between lineages. Concerning the basic structure of cells, some reaction kinetics share the same rates between ODE systems.

In heterogeneous cell population models, the averaged population concentration $[X_{avg}]$ of each reactant X is calculated by the weighted average in Equation (6.1) with population proportion α_i of cell type i. The population proportions are fixed constants with the sum equaling to 1, as shown in Equation (6.2). The reason of the fixed cell proportions is that cell division rates are assumed to be the same between cell types. Therefore, even though the number of cells is increasing due to continuous cell division, the ratio between cell types always remains constant, such as the example shown in Figure 6.1 with the ratio $2:1$ between cell types ■ and ▲, i.e. $\alpha_{■} = 2/3$ and $\alpha_{▲} = 1/3$. Then we have the averaged concentration of X,

$$[X_{avg}] = \sum_i \alpha_i [X_i], \tag{6.1}$$

where

$$\sum_i \alpha_i = 1. \tag{6.2}$$

Example 6.1.1. *Heterogeneous Cell Population*

Consider the following heterogeneous population example consisting of two cell types, ■ *and* ▲*. Let* $X_{■}$ *and* $X_{▲}$ *denote the reactant in* ■ *and* ▲*, respectively.*

$$\xrightarrow{\beta} X_{■} \xrightarrow{\delta}$$
$$\xrightarrow{\beta+\psi} X_{▲} \xrightarrow{\delta}$$

Both of them have their own reaction networks but share the same rates of some reaction kinetics, including birth β *and death* δ*. There exists an extra birth rate* ψ *of cell type* ▲ *to make two networks differently. Then Equations (6.3) and (6.4) are the corresponding linear first-order non-homogeneous ODEs of* ■ *and* ▲ *cell networks that are independent of each others.*

$$\frac{d[X_{■}]}{dt} = -\delta[X_{■}] + \beta \tag{6.3}$$
$$\frac{d[X_{▲}]}{dt} = -\delta[X_{▲}] + \beta + \psi \tag{6.4}$$

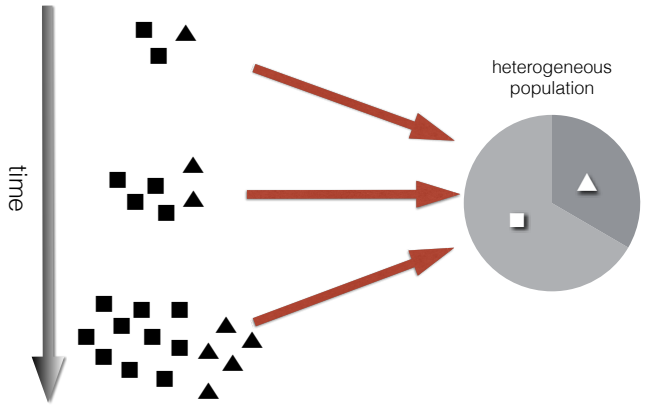

Figure 6.1: A heterogeneous cell population consists of two cell types, ■ and ▲, with population ratio 2 : 1, i.e. population proportions $\alpha_{■} = 1/3$ and $\alpha_{▲} = 2/3$. The population proportions are fixed by the assumption of the same cell division rates between cell types.

Then the analytical explicit solutions are

$$[X_{■}] = [X_{■}]_0 e^{-\delta t} + \frac{\beta}{\delta}(1 - e^{-\delta t}) \tag{6.5}$$

$$[X_{▲}] = [X_{▲}]_0 e^{-\delta t} + \frac{\beta + \psi}{\delta}(1 - e^{-\delta t}). \tag{6.6}$$

We assume the population proportion of cell type ▲ is a fixed constant $\alpha_{▲} =: \alpha$, and hence, the population proportion of cell type ■ is $\alpha_{■} = 1 - \alpha$. Threrefore, the weighted average of reactant X in the heterogeneous population is calculated by substituting the solutions from Equations (6.5) and (6.6) into Equation (6.1).

$$[X_{avg}] = (1 - \alpha)[X_{■}] + \alpha[X_{▲}] \tag{6.7}$$

$$= (1 - \alpha)[X_{■}]_0 e^{-\delta t} + \alpha[X_{▲}]_0 e^{-\delta t} + \frac{\beta + \alpha\psi}{\delta}(1 - e^{-\delta t}) \tag{6.8}$$

If the initial values are the same between cell types, i.e. $[X_{■}]_0 = [X_{▲}]_0 =: [X]_0$, then the averaged concentration becomes simpler.

$$[X_{avg}] = [X]_0 e^{-\delta t} + \frac{\beta + \alpha\psi}{\delta}(1 - e^{-\delta t}) \tag{6.9}$$

This heterogeneous population approach is then applied in the study for the T helper 17 (Th17) cell differentiation. Two cell types, Th0 cell and Th17 cell, have their own GRNs to response the adaptive immune system. Two of three alternative models introduced later, replicate-independent heterogeneous population (M_2) and replicate-dependent heterogeneous population (M_3), are under this heterogeneous cell population structure.

6.2 T Helper Cell Networks

The T helper (Th) cell networks are constructed based on the existing biological knowledge that Th17 cell differentiation is mainly regulated by two transcription factors (TFs), STAT3 and RORγt, in which regulation of RORγt is controlled by the protein product of STAT3 [2]. In the following networks, we let $STAT3_m$ and $STAT3_p$ denote STAT3 in mRNA and protein levels, respectively. Similarly, $RORC_m$ and $RORC_p$ represent RORC in mRNA and protein levels respectively, even though the actual protein name of gene RORC is RORγt. Besides the above four reactants, there is T cell receptor (TCR) activation signal denoted by *Acti* given at the beginning of the experiments. In our cell type specific networks, three main kinds of cellular processes, transcription, translation and degradation, are assumed to be linear reactions.

6.2.1 T Helper 0 Cell Network

(i) Th0 cell dynamics is triggered by the exponential decreasing activation signal *Acti* that up-regulates the transcription process of $STAT3_m$ with the rate γ_1.

$$Acti \xrightarrow{\gamma_1} STAT_m$$

(ii) Besides, $STAT3_m$ is also synthesized with the basal transcription rate β.

$$\xrightarrow{\beta} STAT3_m$$

(iii) Then the TF, $STAT3_p$, is produced by $STAT3_m$ through protein synthesis with the translation rate λ_1.

$$STAT3_m \xrightarrow{\lambda_1} STAT3_p$$

(iv) Further, the mentioned well-known transcription of $RORC_m$ is up-regulated by $STAT3_p$ with the rate γ_2.

$$STAT3_p \xrightarrow{\gamma_2} RORC_m$$

(v) In addition, there is a translation to produce $RORC_m$ with the rate λ_2. However, the last mechanism (v) is not important in Th0 cell network because $RORC_p$ does not directly or indirectly regulate any observed variables, which are $STAT3_m$ and $RORC_m$ in the RNA-Seq data.

$$RORC_m \xrightarrow{\lambda_2} RORC_p$$

Figure 6.2 shows all the above reactions, except the negligible translation of $RORC_p$, in Th0 cell network. Last but not least, there must be degradation rates of all reactants, which are also not shown in Figure 6.2, and they are δ_1, δ_2, δ_3, δ_4 and δ_5 for $Acti$, $STAT3_m$, $STAT_p$, $RORC_m$ and $RORC_p$, respectively. The corresponding coupled linear first-order non-homogeneous ordinary differential equation (ODE) system is

$$\frac{d[Acti]}{dt} = -\delta_1[Acti] \tag{6.10}$$

$$\frac{d[STAT3_m]}{dt} = -\delta_2[STAT3_m] + \gamma_1[Acti] + \beta \tag{6.11}$$

$$\frac{d[STAT3_p]}{dt} = -\delta_3[STAT3_p] + \lambda_1[STAT3_m] \tag{6.12}$$

$$\frac{d[RORC_m]}{dt} = -\delta_4[RORC_m] + \gamma_2[STAT3_p] \tag{6.13}$$

$$\frac{d[RORC_p]}{dt} = -\delta_5[RORC_p] + \lambda_2[RORC_p] \tag{6.14}$$

Further, it can be rewritten in the form of MDE in Equation (6.15)

$$\dot{\mathbf{x}} = \begin{pmatrix} -\delta_1 & 0 & 0 & 0 & 0 \\ \gamma_1 & -\delta_2 & 0 & 0 & 0 \\ 0 & \lambda_1 & -\delta_3 & 0 & 0 \\ 0 & 0 & \gamma_2 & -\delta_4 & 0 \\ 0 & 0 & 0 & \lambda_2 & -\delta_5 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{6.15}$$

where $\mathbf{x} = ([Acti], [STAT3_m], [STAT3_p], [RORC_m], [RORC_p])^T$.

The initial value of activation signal $[Acti]_0$ is 100% at $t = 0$ with a scaling factor S_{Acti}, which is freely chosen and is chosen to be 40 in our application. The initial values of proteins are assumed to be zero because TFs are not expressed before activation. The initial values of mRNAs are directly obtained from RNA-Seq data, in which the averaged values over all replicates are used in M_1 and M_2. Therefore, $\mathbf{x}(0) = (S_{Acti}, [STAT3_m]_0, 0, [RORC_m]_0, 0)^T$.

6.2.2 T Helper 17 Cell Network

Th17 cell network consists of all the reactions in Th0 cell network in addition to two more processes. Note that two networks shares the same rates of the common reaction processes.

(i) One of the additional processes is that the activation signal is boosted by the cytokine inputs. Therefore, there is an extra synthesis of $STAT_m$ in the early

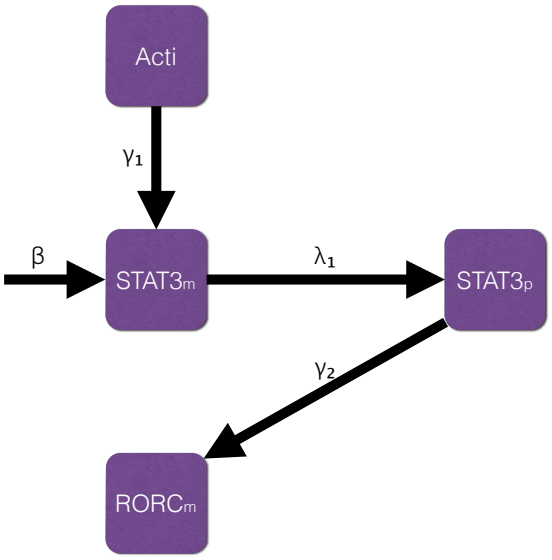

Figure 6.2: The approximative gene regulatory network (GRN) of Th0 cells includes the transcription and translation processes (i)-(iv) as well as the degradation, however, which is not shown in the figure.

period compared to Th0 cells and the transcription rate becomes $(1 + \eta)\gamma_1$ in Th17 cell network.

$$Acti \xrightarrow{(1+\eta)\gamma_1} STAT_m$$

(vi) The second additional cellular process, which is the main part making two networks producing different dynamics, is the feedback transcription to synthesize $RORC_m$ by its protein product $RORC_p$ with the rate ψ.

$$RORC_p \xrightarrow{\psi} RORC_m$$

This creates a loop between $RORC_m$ and $RORC_p$ to have self-regulation within the gene. The corresponding reaction network of Th17 cells is shown in Figure 6.3. Then the corresponding coupled linear first-order non-homogeneous ODE system

is

$$\frac{d[Acti]}{dt} = -\delta_1[Acti] \tag{6.16}$$

$$\frac{d[STAT3_m]}{dt} = -\delta_2[STAT3_m] + (1+\eta)\gamma_1[Acti] + \beta \tag{6.17}$$

$$\frac{d[STAT3_p]}{dt} = -\delta_3[STAT3_p] + \lambda_1[STAT3_m] \tag{6.18}$$

$$\frac{d[RORC_m]}{dt} = -\delta_4[RORC_m] + \gamma_2[STAT3_p] + \psi[RORC_p] \tag{6.19}$$

$$\frac{d[RORC_p]}{dt} = -\delta_5[RORC_p] + \lambda_2[RORC_m] \tag{6.20}$$

Further, it can be rewritten in the form of MDE in Equation (6.21)

$$\dot{\mathbf{x}} = \begin{pmatrix} -\delta_1 & 0 & 0 & 0 & 0 \\ (1+\eta)\gamma_1 & -\delta_2 & 0 & 0 & 0 \\ 0 & \lambda_1 & -\delta_3 & 0 & 0 \\ 0 & 0 & \gamma_2 & -\delta_4 & \psi \\ 0 & 0 & 0 & \lambda_2 & -\delta_5 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 0 \\ \beta \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{6.21}$$

where $\mathbf{x} = ([Acti], [STAT3_m], [STAT3_p], [RORC_m], [RORC_p])^T$ and
$\mathbf{x}(0) = (S_{Acti}, [STAT3_m]_0, 0, [RORC_m]_0, 0)^T$, which shares the same as in Th0 cell network.

6.3 Alternative Population Models

In this section, we use both homogeneous and heterogeneous population approaches to model the dynamical system in cellular differentiation. Alternative models are often generated to represent dynamical systems given the same data set. Due to uncertainties of the true system, several choices are able to describe the same observed data, however, with different errors.

In the RNA-Seq data of the Th17 cell differentiation, there are two experiments, controlled and variable. Remind that naïve $CD4^+$ T (Thp) cells require a T cell receptor (TCR) activation under a particular cytokine environment to differentiate. Therefore, Thp cells are given an activation in the presence of cytokines, IL-1b, IL-6 and low concentration of TGF-β, in the variable experiment, which is the experiment aiming to produce Th17 cells. On the other hand, there is no cytokines given in the controlled experiment that the cells are supposed to remain as activated Th (Th0) cells.

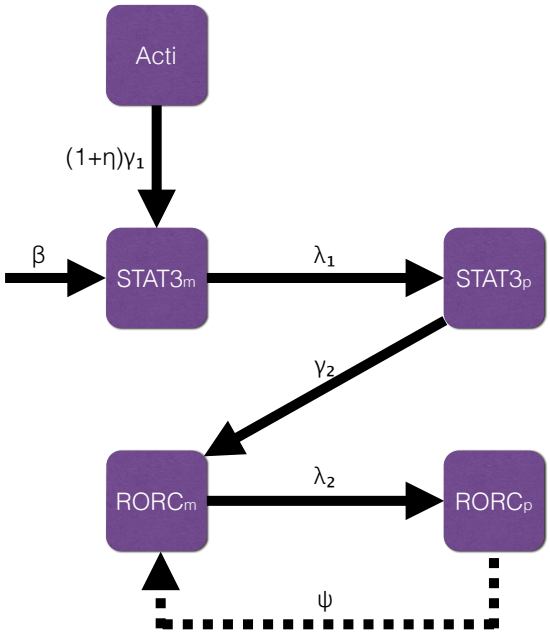

Figure 6.3: The approximative gene regulatory network (GRN) of Th17 cells includes the transcription and translation processes (i)-(vi) as well as the degradation, however, which is not shown in the figure.

Three following alternative models, homogeneous population (M_1), replicate-independent heterogeneous population (M_2) and replicate-dependent heterogeneous population (M_3), describe two experiments in different population approaches. Table 6.1 shows the cell population structures in three alternative models and heterogeneity is only considered in the variable experiment only.

	Model	Cont.	Vari.
Homogeneous population	M_1	Th0	Th17
Replicate-independent heterogeneous population	M_2	Th0	Th0+Th17
Replicate-dependent heterogeneous population	M_3	Th0	Th0+Th17

Table 6.1: Three alternative population models consist of different cell types in the experiments. All three models consist of pure Th0 cells in the controlled experiment. In the variable experiment, M_1 consists of pure Th17 cells, while M_2 and M_3 consist of both Th0 and Th17 cells.

6.3.1 M_1 - Homogeneous Population

The homogeneous population model (M_1) is a pure population structure under a usual modeling approach. In the cell differentiation experiment, Thp cells are assumed to differentiate into only one lineage, which is Th0 cells in the controlled experiment and Th17 cells in the variable experiment, as shown in Table 6.1. Therefore, M_1 uses Th0 and Th17 cell networks with their ODE systems individually to describe the controlled and variable experiments, respectively.

6.3.2 M_2 - Replicate-independent Heterogeneous Population

The replicate-independent heterogeneous population model (M_2) takes the heterogeneity into account in the variable experiment, whereas the pure Th0 cell population is assumed to be the same as M_1 in the controlled experiment. Biologists found that Thp cells may not fully differentiate and even differentiate into several lineages to form a heterogeneous cell population [3, 8, 9, 38]. In the variable experiment, a portion of cells is assumed to remain as Th0 cells due to unknown reasons such as signal losing, and therefore, the dynamics is represented by the subpopulation of Th0 and Th17 cells in parallel.

We let $\alpha := \alpha_{Th17}$ denote the fixed population proportion of Th17 cells, and then, let $\alpha_{Th0} = 1 - \alpha$ denote the one of Th0 cells. The averaged concentration of reactants X, including $Acti$, $STAT3_m$, $STAT_p$, $RORC_m$ and $RORC_p$, are calculated by Equation (6.1) becoming

$$[X_{avg}] = \alpha_{Th17}[X_{Th17}] + \alpha_{Th0}[X_{Th0}] \tag{6.22}$$
$$= \alpha[X_{Th17}] + (1-\alpha)[X_{Th0}], \tag{6.23}$$

where $[X_{Th17}]$ and $[X_{Th0}]$ are the concentration of X in Th17 and Th0 cells, respectively. Figure 6.4 shows the difference between the homogeneous and heterogeneous cell population in Th17 cell differentiation.

6.3.3 M_3 - Replicate-dependent Heterogeneous Population

The replicate-dependent heterogeneous population (M_3) takes the heterogeneity into account as M_2, however, with replicate dependence. Experiments usually are repeated by replicates, three replicates in our data, and each may not produce exactly the same dynamics. Therefore, each replicate should have its individual dynamics.

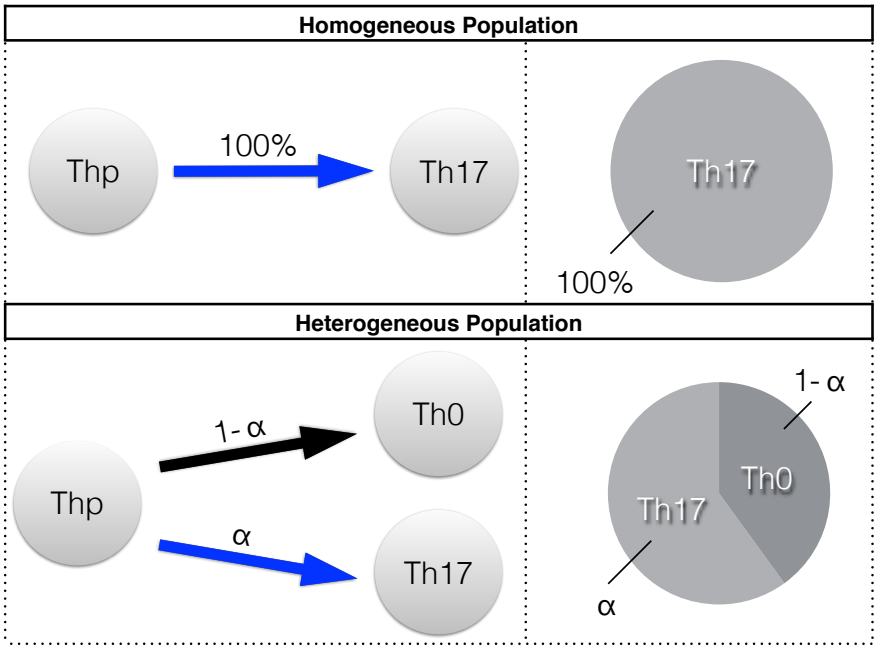

Figure 6.4: The upper part shows that 100% of Thp cells is assumed to differentiate into Th17 cells in the homogeneous population. The lower part shows that only a portion α of Thp cells is assumed to differentiate into Th17 cells but the rest remain as Th0 cells during the cell differentiation in the heterogeneous population.

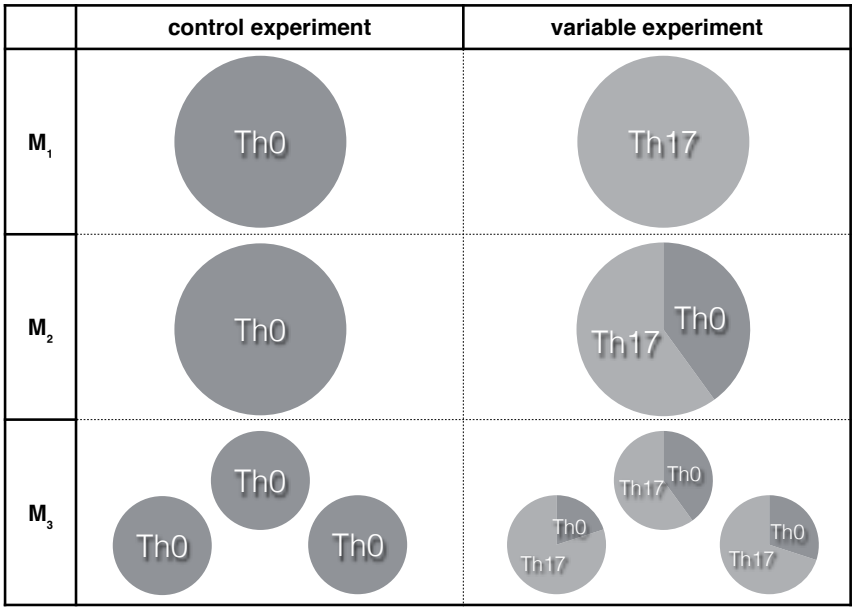

Figure 6.5: All models describe pure Th0 cell population in the controlled experiment, while M_1 is a homogeneous population model but M_2 and M_3 are heterogeneous population models to describe Th17 cell population in the variable experiment. On the other hand, M_1 and M_2 are replicate-independent models that consider the same dynamics between replicates, whereas M_3 is replicate-dependent model with three independent population proportions of Th17 cells, α_1, α_2 and α_3, in the variable experiment.

In this model, replicates are assumed to have independent differentiation proportions α_r and own initial values while all reaction rates are shared. The main difference depends on the population proportions, α_1, α_2 and α_3, and the initial values only give a tiny shift between replicates. Figure 6.5 shows that each replicate in M_3 is different and with its own cell population in the variable experiment, while other models, M_1 and M_2, share the same dynamics between replicates within the experiment.

In the Th17 cell differentiation application, M_3 produces three different responses because of the population proportions and initial values in the variable experiment but three similar responses because of the initial values only in the controlled experiment.

6.4 Efficient Evaluation of Explicit Solutions

Given the gene regulatory network (GRN) of Th0 and Th17 cells in the form of matrix differential equations (MDEs) in Equations (6.15) and (6.21), we are able to solve them using the explicit general solution in Equation (3.16) of the system in Equation (3.13) instead of using numerical ODE solvers. The analytical solutions are available by solving the matrix exponential $e^{\mathbf{A}t}$ analytically. On one hand, we can substitute the proposed parameter set θ, including the coefficients of the MDE system and the population proportions, into the analytical solution $\mathbf{x}(\theta, t)$ at each iteration during sampling. On the other hand, we can first substitute the proposed parameter set θ into the MDE system, and then solve the matrix exponential $e^{\mathbf{A}t}$ numerically to obtain the numerical explicit solution $\mathbf{x}(t)$. In our application, the response of Th0 cell network is obtained by the former way and the latter way is used to get the response of Th17 cell network. The reason of using numerical method in Th17 cell network is that the analytical solution $\mathbf{x}(\theta, t)$ of Equation (6.21) is too complicated due to the loop structure within the self-regulated gene RORC. Consequently, substitution into the solution is more expensive, and hence, solving the matrix exponential $e^{\mathbf{A}t}$ numerically is a better choice.

In practice, data likelihood needs only the response at some time points, to compare with the data set. The matrix exponential $e^{\mathbf{A}t}$ is chosen to be solved by the scaling and squaring method [27] directly using the built-in function `expm` in MATLAB. It can be simply calculated by `expm(A*t)` at each time point t or by `expm(A)^t`, which reflects the property $e^{\mathbf{A}t} = (e^{\mathbf{A}})^t$. The latter way taking the resulting number to the power t can reduce the number of calling the time-consuming function `expm` to make computation faster. However, there is a significant error while $t \notin \mathbb{N}$, and therefore, we use `expm(A*t)` for $t \notin \mathbb{N}$ and `expm(A)^t` for $t \in \mathbb{N}$.

6.5 Statistical Models of RNA Sequencing Data

In order to combine mathematical modeling with discrete read count RNA sequencing (RNA-Seq) data, we use the framework based on a discrete probability distribution called the negative binomial distribution, or also named gamma-Poisson distribution. It is a popular alternative statistical model other than Poisson model, in which its mean and variance are the same, of count data in numerous biological applications [12, 13, 14, 15, 16]. Since in situations where there is a positive correlation in the occurrence of events, the observed variation is significantly greater than the mean. The negative binomial distribution therefore provides a better ap-

proximation of data, such as in serial analysis of gene expression. The probability mass function (PMF) of the negative binomial distribution is

$$\mathcal{L}(\mu, \phi|y) = \frac{\Gamma(y + \phi^{-1})}{\Gamma(\phi^{-1})\Gamma(y+1)}(\frac{1}{1+\mu\phi})^{\phi^{-1}}(\frac{\mu}{\phi^{-1}+\mu})^{y}, \tag{6.24}$$

where Γ is the gamma function, μ is the mean and ϕ is the dispersion [12, 13].

Let Y be a negative binomial random variable with mean μ and dispersion ϕ, denoted $Y \sim \mathcal{NB}(\mu, \phi)$. Then $\mathbb{E}(Y) = \mu$ and $var(Y) = \mu + \phi\mu^2$. Since the dispersion $\phi > 0$ by definition, the negative binomial distribution has an over-dispersion property, i.e. $var(Y) > \mathbb{E}(Y)$, to express the positive correlation events.

The human experimental data was measured in RNA-Seq technique. As mentioned, there are two experiments, controlled and variable. Each aligned data set contains three-replicate measurements of all the known genes at ten time points $\mathbf{t} = \{0, 0.5, 1, 2, 4, 6, 12, 24, 48, 72\}$ (in hour [h]), in which initial values are fixed in the model and therefore not included in the likelihood. In fact, we need the read count data of only two genes, STAT3 and RORC. Hence, there are in total $2 \times 3 \times 9 \times 2 = 108$ independent data points y_{grns} representing the discrete points of gene g at the n^{th} time point t_n in the r^{th} replicate for the experiment s, where g is STAT3 and RORC (denoted by 1 and 2 respectively), $r = 1, 2, 3$, $n = 2, 3, ..., 10$ and s is controlled and variable condition (denoted by 1 and 2 respectively).

The data likelihood function is assumed to be negative binomial distributed to model the data set $D = \{\mathbf{y}, \mathbf{t}\}$ with the mean $\boldsymbol{\mu}$ and dispersion $\boldsymbol{\phi}$. There are several ways to determine the dispersions depending genes and time. Figure 6.6 shows the dispersions that are time-dependent values ϕ_n estimated over all the genes and gene-time-dependent values ϕ_{gn} as well as their time-averaged values, $\bar{\phi}$ and $\bar{\phi}_g$. Note that all dispersions are estimated using edgeR package [14] directly from the measurements as a pre-processing that considers pairing between both controlled and variable experiments as well as all replicates. Due to the similarity of the estimated dispersions between genes, the data likelihood uses the time-dependent dispersions ϕ_n over all the genes and the predictive likelihood needs their time-average $\bar{\phi}$ to predict dynamics at the continuos time.

Hence, we can write $y_{grns} \sim \mathcal{NB}(\mu_{grns}, \phi_n)$ and the data likelihood can be defined as

$$p(D|\boldsymbol{\theta}) = \prod_{g=1}^{2}\prod_{r=1}^{3}\prod_{n=2}^{10}\prod_{s=1}^{2}\mathcal{L}(\mu_{grns}(\boldsymbol{\theta}), \phi_n|y_{grns}), \tag{6.25}$$

where $\mathcal{L}(\mu, \sigma|y)$ is the probability mass function (PMF) defined in Equation (6.24). Remind that the explicit solution $\mathbf{x}(\theta, t)$ of mRNA represents the relative mRNA

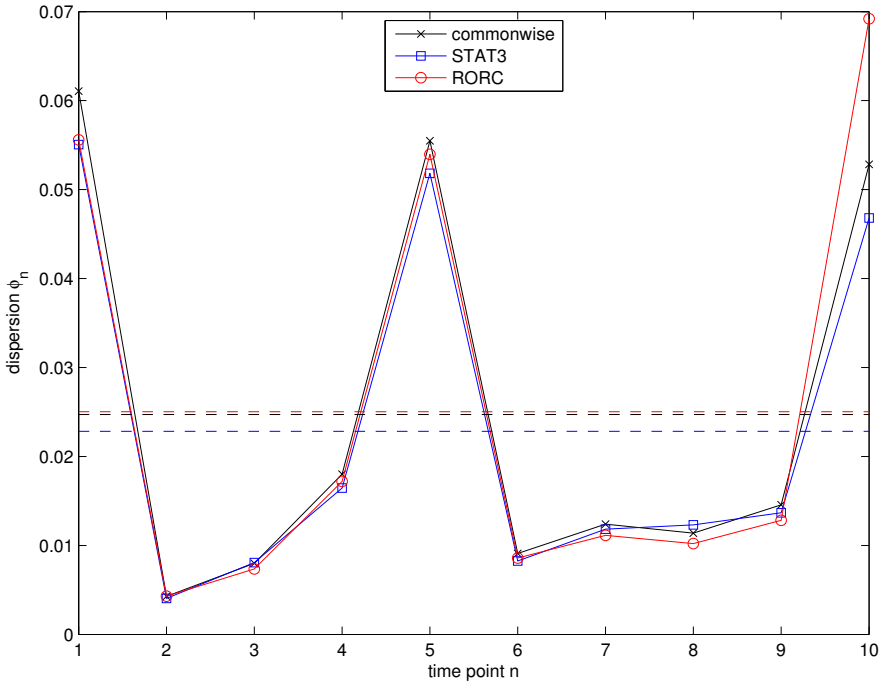

Figure 6.6: The figure shows dispersion values defined in the negative binomial distribution. The black solid line shows the time-dependent values ϕ_n estimated over all the genes, while the blue and red solid lines show the gene-time-dependent values ϕ_{gn} of genes STAT3 and RORC, respectively. The time-averaged values, $\bar{\phi}$ and $\bar{\phi}_g$, are shown the dashed lines in the corresponding colors. Note that all dispersions are estimated directly from the measurements as a pre-processing [14].

abundance, which can be considered as the normalized RPKM values defined in Equation (2.1). Therefore, the mean can be expressed in the form $\mu_{grns}(\boldsymbol{\theta}) = x_{gns}(\boldsymbol{\theta})N_{rns}L_g \cdot 10^{-9}$ in the replicate-independent models M_1 and M_2, and $\mu_{grns}(\boldsymbol{\theta}) = x_{grns}(\boldsymbol{\theta})N_{rns}L_g \cdot 10^{-9}$ in the replicate-dependent model M_3, where N_{nrs} is the library size of sequencing snapshot, L_g is the exon length and mRNA responses $x_{gns}(\boldsymbol{\theta})$ and $x_{grns}(\boldsymbol{\theta})$ are in RPKM values.

On the other hand, the predictive likelihood in Equation (4.16) can be defined similarly to the data likelihood, but using the time-averaged dispersion parameter $\bar{\phi}$ from the original estimation as well as the time-averaged library size $\bar{N}_{rs}$ in replicate-dependent model (M_3) and time-replicate-average $\bar{N}_s$ in replicate-independent models (M_1, M_2).

For prior, we are interested in the coefficients representing reaction rates in the systems and the cell differentiation population proportions. Two coefficients,

λ_1 and λ_2, are fixed due to the dependence between other coefficients in the explicit solutions of observables, mRNAs. Besides, the initial values and dispersions are also fixed. Then there are ten coefficients denoted by a set $\tilde{\boldsymbol{\theta}} = \{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \gamma_1, \gamma_2, \beta, \psi, \eta\}$ and the population proportion $\boldsymbol{\alpha}$ are sampled, and hence, the parameter dimension d are 10, 11 and 13 for M_1, M_2 and M_3, respectively. The uninformative prior is assumed to be standard log-normal distributed for coefficient $\tilde{\theta}_m$ and uniformly distributed between 0 and 1 for population proportion α_i, where $m = 1, 2, ..., 10$, $i = 1$ in M_2 and $i = 1, 2, 3$ in M_3. Therefore, the parameter prior is defined as

$$p(\boldsymbol{\theta}) = \prod_{m=1}^{10} \rho_N(\tilde{\theta}_m) \tag{6.26}$$

in M_1 and

$$p(\boldsymbol{\theta}) = \prod_{m=1}^{10} \rho_N(\tilde{\theta}_m) \prod_{i=1}^{d-10} \rho_U(\alpha_i) \tag{6.27}$$

in M_2 and M_3, where $\rho_N(\theta)$ is the probability distribution function (PDF) of the log-normal distribution defined in Equation (9.3) and $\rho_U(\theta)$ is the PDF of the uniform distribution between 0 and 1 defined in Equation (9.5).

In practice, the log-likelihood function, which is the natural logarithm of the likelihood function, is more convenient to use. The logarithm of a function has its maximum at the same points as the function itself because of monotonicity of the logarithm. The likelihood can be replaced by log-likelihood in maximum likelihood estimation and even distribution estimation. It is also easier to handle than the original function because the logarithm of a product is a sum of individual logarithms.

$$\log \prod_{n=1}^{N} \mathcal{L}(\theta|y_n) = \sum_{n=1}^{N} \log \mathcal{L}(\theta|y_n) \tag{6.28}$$

For the same reason, the log-prior function is also convenient to use in the actual computational steps.

$$\log \prod_{m=1}^{d} \rho(\theta_m) = \sum_{m=1}^{d} \log \rho(\theta_m) \tag{6.29}$$

6.6 Computational Implementation

The computation of mathematical models, sampling MCMC algorithms and marginal likelihood estimation were implemented in MATLAB. The matrix exponentials as

mentioned were computed using the built-in function `expm` [27]. All the sampling results were based on popMCMC algorithm with 5 independent chains each of $J = 2 \times 10^6$ iterations in 10 temperatures without rejection of chains due to good convergence. During the burn-in period, which is the first half of each chain (i.e. 10^6 iterations), the proposal distributions were tuned adaptively. Precisely, each subchain with $J_s = 500$ sub-iterations applies adaptation 2000 times updating the proposal covariance matrices Σ_{τ_n} from initial proposal covariance matrices defined in Equation (6.30)

$$\Sigma_{\tau_n} = w_{\tau_n} \mathbf{I}_d, \tag{6.30}$$

where $w_{\tau_n} = 0.1 \times \left[1 - \frac{(n-1)^3}{(10/\sigma_s)^3 + (n-1)^3}\right]$ is a weight factor, $\mathbf{I}_d$ is an identity matrix of size d and σ_s is a shape parameter that was chosen to be 6. On the other hand, the temperature schedule, which is employed in thermodynamic integration for the power posterior estimator, is defined following the article [20] as

$$\tau_n = \left(\frac{n-1}{10-1}\right)^5. \tag{6.31}$$

As mentioned in Section 5.4, the resulting samples have to go through the processes, burn-in and thinning, to discard first half samples of the chains and then to save every 1000^{th} iterations to obtain 5000 low autocorrelation samples from in total 10^7 samples.

Chapter 7

Results

In this chapter, we show the results of the T helper 17 (Th17) cell differentiation from both simulated and experimental data. The analysis mainly focuses on estimation of parameter posterior distributions and model ranking to compare alternative models given observed data. The results show that the models are well-designed, parameters are identifiable and true models are discovered to show model reliability by calculating the marginal likelihoods. The key result is that M_3 is top-ranked with the strongest evidence given the experimental data. Further, the molecular dynamics of Th cell population can be predicted by the dominant model M_3.

7.1 Simulated Data

Before drawing conclusions from the experimental data, we can repeat the analysis with simulated data sets to verify the model reliability. A set of true parameters is freely chosen to simulate a data set, and then the corresponding parameter posterior distribution based on this data set should have a high density near the true parameters to verify the estimation. Moreover, we can calculate marginal likelihood of each model to show the true model having the strongest evidence.

7.1.1 Data Generation

In the simulation, there are three data sets (D_{M_1}, D_{M_2}, D_{M_3}) generated from all three alternative models. First, the true parameters are chosen to be the following. $\tilde{\boldsymbol{\theta}} = \{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \gamma_1, \gamma_2, \beta, \psi, \eta\} = \{0.5, 0.5, 0.5, 0.5, 0.2, 0.5, 0.01, 10, 0.09, 2\}$ for

all models, $\alpha = 0.4$ in M_2 and $\boldsymbol{\alpha} = \{0.4, 0.5, 0.6\}$ in M_3. Then there are true responses of ODE systems. Hence, using the data likelihood function, errors can be added randomly according to the dispersions ϕ_n at discrete time points.

7.1.2 Identifiable Parameter Posterior Distributions

From all three simulated data sets, we conclude that the estimation of parameter posterior distributions works perfectly and all the parameters are identifiable. The parameter posterior distributions are significantly updated from the uninformative prior distributions and all the independent chains converge to the same equilibrium distribution as well as explore the true parameters in the simulated data. More importantly, the estimated potential scale reduction $\hat{R}$ are close to 1 for all parameters that means the samples are well mixed and show mathematically converge to one equilibrium distribution. The figures of the estimated potential scale reduction can be found in the appendix, Figure 9.1.

Note that the resulting estimation of parameter posterior distributions from data sets D_{M_1}, D_{M_2} and D_{M_3} are single-mode distributed. However, there is possibility that parameter posterior distributions are multi-mode distributed, depending on randomness of generating data sets. For example, parameter posteriors are two-mode distributed given the simulated data set D'_{M_1}. The figure can be found in the appendix, Figure 9.3.

7.1.3 Model Ranking

The model ranking given simulated data shows reliable results to verify model reliability by ranking the original true models correctly. The true models, which are the models simulating data, are ranked correctly with the highest model posteriors for all three alternative models. Precisely, the natural logarithm of the marginal likelihoods and model posteriors are calculated by the power posterior estimators, as shown in Table 7.1. All the true models are ranked with the highest model posteriors 63%, 94% and 100% given D_{M_1}, D_{M_2} and D_{M_3}, respectively. These results are good to support the model reliability, especially of M_3, for further extension to the experimental data.

	M_1	M_2	M_3
D_{M_1}	**-807.55 (63%)**	-808.08 (37%)	-823.75 (0%)
D_{M_2}	-788.82 (6%)	**-785.99 (94%)**	-795.00 (0%)
D_{M_3}	-813.07 (0%)	-812.22 (0%)	**-795.84 (100%)**
D_E	-1403.0 (0%)	-1405.6 (0%)	**-1394.5 (100%)**

Table 7.1: The natural logarithm of the marginal likelihoods are estimated by the power posterior estimator for all three alternative models. The corresponding model posteriors are shown in brackets. There are in total four data sets, including three simulated data sets D_{M_1}, D_{M_2} and D_{M_3} as well as one experimental data set D_E.

7.2 Experimental Data

After verifying model reliability from simulated data, we can repeat the analysis on the experimental data and draw conclusions in the real application. From the estimation of parameter posterior distributions and model ranking, the results decisively favor M_3. Further, we predict the molecular dynamics using the predictive likelihood with the dominant model M_3.

7.2.1 Estimated Parameter Posterior Distributions

The estimation of parameter posterior distributions given the experimental data D_E works perfectly as the simulated data. The parameter posterior distributions are single-mode distributed and significantly updated from the prior, as shown in the appendix, Figures 9.4, 9.5 and 9.6. All the independent chains converge to the same equilibrium distribution with the estimated potential scale reduction $\hat{R}$ near to 1 that can be found in the appendix, Figure 9.2.

More importantly, the estimated population proportion α_i that involves in the variable experiment has interesting results. Figure 7.1 shows the estimated marginal distribution of population proportions in M_2 and M_3. The marginal distribution of α for M_2 and the marginal distribution of α_2 for M_3 are mainly distributed at the upper bound $\alpha = 1$. This means that all replicates in M_2 are estimated to have a high probability of pure Th17 cell population, which is the assumption of M_1. This result shows that M_1 and M_2 have the similar cell population structure given the data. On the other hand, the similar distribution of α_2 in M_3 explains that the 2^{nd} replicate dominates the population proportion in M_2. Therefore, M_3 is able to show the different population proportions between replicates that the cell populations of 1^{st} and 3^{rd} replicates do not fully differentiate into Th17 cells,

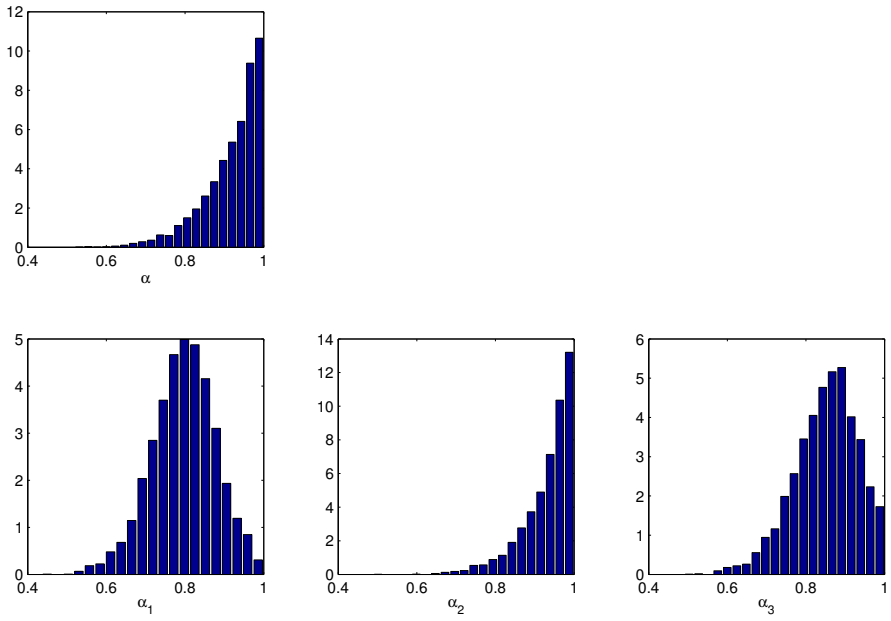

Figure 7.1: The figures show the estimated marginal distribution of population proportions in M_2 and M_3. The upper figure shows the marginal parameter posterior distribution of α in M_2. The lower figures show the marginal parameter posterior distributions of α_1, α_2 and α_3 in M_3. Note that α and α_2 are mainly distributed at the upper bound, 100%.

while the 2^{nd} replicate has a high probability that most of the cell population are Th17 cells. Before ranking the alternative models, we can already guess that M_3 should model the data the best, and M_1 and M_2 have similar model evidences. The following section shows this is true quantitatively by marginal likelihoods.

7.2.2 Model Ranking

The same analysis was applied on the experimental data showing the strongest model evidence for M_3. Similar to the results obtained from the data D_{M_3}, M_3 is ranked the top with 100% model posterior, however, with lower marginal likelihood value from D_E, as shown in Table 7.1. This mathematically proves that M_3 is decisively better to model the data to describe the differences between replicates.

To be more specific, we can use Bayes factor to show detailed comparison between

models. The Bayes factor in log-scale written in the matrix form is

$$\log \mathbf{B} = \begin{pmatrix} 0 & 2.6 & -8.5 \\ -2.6 & 0 & -11.0 \\ 8.5 & 11.0 & 0 \end{pmatrix}. \tag{7.1}$$

The resulting values $\log B_{31} = 8.5$ and $\log B_{32} = 11.0$ express the same result as $p(M_3|D_E) = 100\%$ supporting M_3 to be the decisively favored model. Another resulting value $\log B_{12} = 2.6$ illustrates the comparison that model posteriors cannot show, which is that M_1 is strongly favored than M_2, while both model posteriors of M_1 and M_2 are 0% with the consideration of M_3.

7.2.3 Posterior Predictive Distributions

Given the dominant model M_3 as the result from model ranking, the molecular dynamics can be predicted using Equations (4.16) and (4.17) with the predictive likelihood defined before. The dynamics of two genes, STAT3 and RORC, in mRNA levels can be shown by the marginal posterior predictive distributions with measurement uncertainties, which are represented by dispersions in the likelihood. The predictive dynamics is capable of modeling the observed data. Figure 7.2 shows the replicate-dependent predictions in both controlled and variable experiments with the experimental data. On the other hand, we can also predict the unobserved variables including two transcription factors (TFs) $STAT3_p$ and $RORC_p$ as well as the activation signal $Acti$ in the networks. The predictions of unobserved variables are generated using parameter uncertainties only and therefore provide a smaller variation of the distributions. The marginal posterior predictive distributions of all variables can be found in the appendix, Figure 9.7.

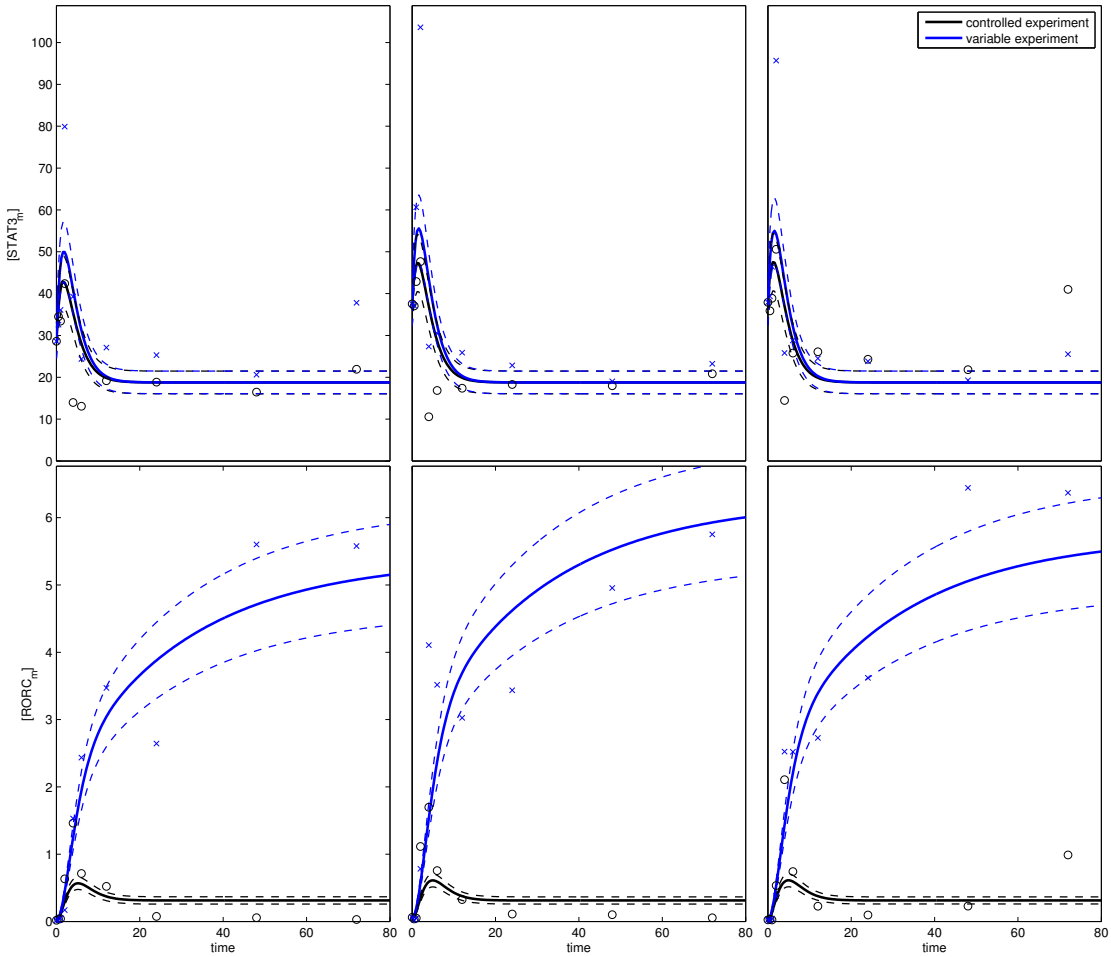

Figure 7.2: Posterior predictive distributions are generated using the dominant model, M_3. The figures on the 1^{st} and 2^{nd} row show time-dependent marginal predictive distributions of $[STAT3_m]$ and $[RORC_m]$ respectively, using the mean (solid lines) and one standard deviation interval (dashed lines). The lines are plotted using black and blue color for controlled and variable experiments, respectively. The normalized RPKM data are plotted using black circles and blue cross marks for controlled and variable experiments, respectively. Both predictions and data are replicate-dependent, shown in the 1^{st}, 2^{nd} and 3^{rd} columns representing 1^{st}, 2^{nd} and 3^{rd} replicates, respectively.

Chapter 8

Discussion and Conclusions

In summary, we presented modeling approaches taking homogeneous and heterogeneous populations into account and compared three alternative models for Th17 cell differentiation using Bayesian inference with MCMC sampling methods. Precisely, we used ODE models written in the form of MDEs that have general solutions to describe molecular dynamics of cells and then used statistical methodology to draw conclusions based on the experimental data. In Bayesian computation, we used popMCMC algorithm to draw samples from target distributions including both single-mode and multi-mode distributions. Moreover, we used the power posterior estimators on marginal likelihood calculations which show reliable results. Most importantly, the estimation provided the strongest evidence for the replicate-dependent heterogeneous population model (M_3) to show the possibility of heterogeneity in the Th17 cell differentiation.

The main idea of the study is to present the novel idea to model heterogeneity in cell differentiation processes. Essentially, we are able to show that the heterogeneous population models (M_2, M_3) are reliable by model ranking from simulated data as well as M_3 is top-ranked from the experimental data. The model ranking from simulated data obviously shows that the homogeneous population model (M_1) is not preferred to describe the data D_{M_2} and D_{M_3} generated from the heterogeneous population. However, M_2 is able to describe the data D_{M_1} generated from M_1 with 37% model posterior. In other words, heterogeneous population approach is capable of modeling homogeneous population structures but not vice versa. On the other hand, the model ranking from the experimental data D_E shows that two of the three replicates apparently consists of Th0 and Th17 cells while the one replicate has a high possibility that consists of pure Th17 cells in the variable experiment. Nevertheless, without the consideration of replicate dependence (i.e. M_3), Bayes factor $\log B_{12} = 3$ shows that the homogeneous population

model (M_1) is strongly preferred than the heterogeneous population model (M_2). Anyway, the heterogeneity should be taken in the consideration in mathematical modeling while there is a possibility of existing different subpopulations. The approach in this study can be applied generally in other applications, for example, the iTreg cell differentiation introduced in the following. Previously, Hasenauer et al. [38] have developed a method to analyze heterogeneous populations using ODE constrained mixture models to study NGF-induced Erk1/2 phosphorylation in primary sensory neurones, a process relevant in inflammatory and neuropathic pain. However, they focus on the single-cell dynamics that captures cell-to-cell variability with subpopulations. This study has some parallels with our model development but the modeling setting is in general different. More details of this heterogeneous population approach for individual cells can be found in [38].

In the future, there are several ways to improve and extend the modeling. First, the sampling can be extended into a higher parameter dimension, which means that we can learn from more parameters such as initial values and dispersions. The initial values are fixed according the experimental data and dispersions are directly measured as a pre-processing in our models. Raue et al. [39] have compared methods of estimating the measurement noise, including pre-processing and simultaneous estimation. We can follow the same idea to compare whether pre-processing or simultaneous estimation is better in our application. Second, this novel heterogeneous population approach can be easily extended to cover another application, the iTreg cell differentiation. The analysis of Th17 cells can be simply replaced by iTreg cells to form two subpopulations of Th0 and iTreg cells. However, it requires a difficult but necessary task, which is to construct a GRN of iTreg cells based on the existing biological knowledge to model the regulation of the key TF of iTreg cells, FOXP3. Given experimental data of the iTreg cell differentiation, the same method should show the results of the model dynamics and population structure. In addition, the approach can be also extended to cover the reciprocal differentiation of Th17 and iTreg cells with three subpopulations, including Th0, Th17 and iTreg cells. This is a more complicated model requiring the GRNs of three cell types regulating three TFs, STAT3, RORC and FOXP3. To keep the simplest modeling, we can split the networks into two that one regulates STAT3 and RORC as presented in this study and one regulates FOXP3, separately. This may disconnect the inhabitation between RORC and FOXP3 but the model can be easily extended to approximate the dynamics using the same GRNs of the Th17 cell differentiation in this thesis.

Chapter 9

Appendix

9.1 Probability Distributions

9.1.1 Normal Distribution

The normal distribution, or Gaussian distribution, is a continuous probability distribution common used in statistics. It is often a good approximation because of the central limit theorem. This theorem states that the mean of any set of variates with any distribution having a finite mean and variance tends to the normal distribution.

Let Y be a normally distributed random variable with mean $\mathbb{E}(Y) = \mu$ and variance $var(Y) = \sigma^2$, i.e. standard deviation σ, denoted by $Y \sim \mathcal{N}(\mu, \sigma^2)$. If $\mu = 0, \sigma = 1$, i.e. $Y \sim \mathcal{N}(0, 1)$, then the distribution is called the standard normal distribution. It is clear that the normal distribution is of two parameters μ and σ, then the likelihood function is a function of $\theta = \{\mu, \sigma\}$ given y. The probability distribution function (PDF) of the normal distribution is

$$\mathcal{L}(\mu, \sigma|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}. \tag{9.1}$$

The log-likelihood function is more often used than the likelihood function. Then Equation (9.1) becomes

$$\log \mathcal{L}(\mu, \sigma|y) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y-\mu)^2}{2\sigma^2}. \tag{9.2}$$

9.1.2 Log-normal Distribution

The log-normal distribution is the natural logarithm of the normal distribution. It is a suitable distribution of positive defined valuables because its range is from zero to infinity instead of covering the whole axis. The PDF of the log-normal distribution is not only simple taking the logarithm, but also multiplies the Jacobian of the transformation, which is $\frac{1}{\theta}$. The PDF of the log-normal distribution is

$$\rho(\theta) = \frac{1}{\theta}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}}e^{-\frac{(\log\theta-\tilde{\mu})^2}{2\tilde{\sigma}^2}}, \tag{9.3}$$

where $\tilde{\mu}$ is the mean and $\tilde{\sigma}$ is the standard deviation. The log-normal distribution is denoted by $\log\mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ similarly to the normal distribution.

On the other hand, we let Θ be $\log\theta$, i.e. $\Theta := \log\theta$, and then Θ follows the normal distribution with the PDF

$$\rho(\Theta) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}}e^{-\frac{(\Theta-\tilde{\mu})^2}{2\tilde{\sigma}^2}}, \tag{9.4}$$

with the same mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$. This shows the relation between the normal distribution and log-normal distribution.

9.1.3 Uniform Distribution

A uniform distribution, or rectangular distribution, is a probability distribution that has constant probability. It is a simple distribution for both continuous and discrete random valuables. Depending on the applications, continuous or discrete distribution is used. In our application, a continuous uniform distribution is properly defined for parameters such as the population proportions. The PDF of the continuous uniform distribution is

$$\rho(\theta) = \frac{1}{b-a}, \tag{9.5}$$

if $\theta \in [a, b]$, where a and b are the lower and upper bounds respectively, and otherwise equals to zero. The uniform distribution is denoted by $\mathcal{U}[a, b]$.

9.2 Supplementary Figures

9.2.1 Potential Scale Reduction

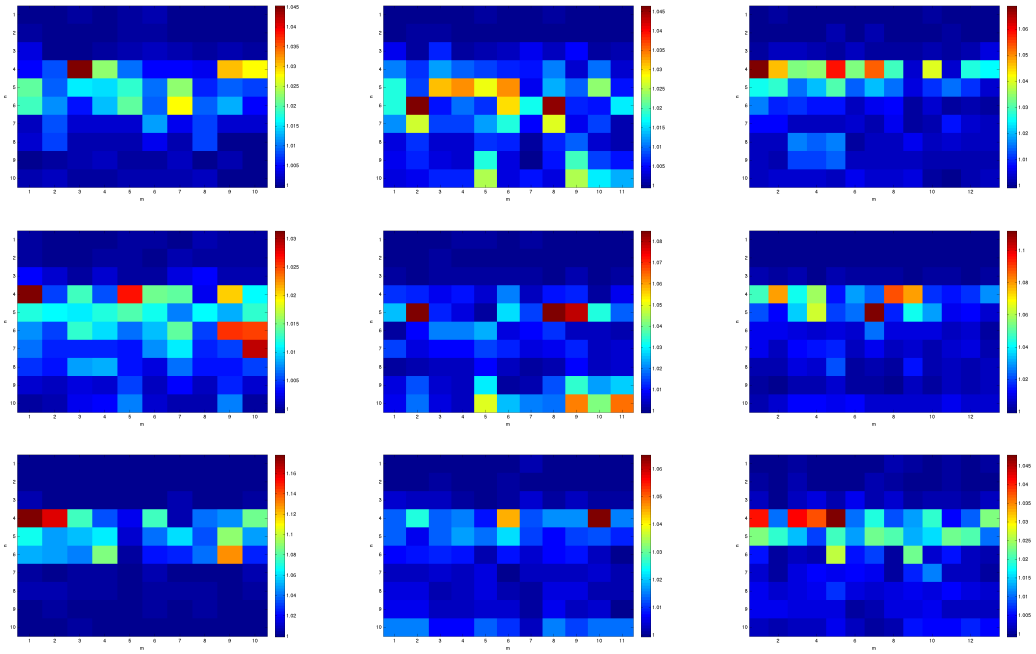

Figure 9.1: The estimated potential scale reduction $\hat{R}$ from simulated data. The figures on the 1^{st}, 2^{nd} and 3^{rd} rows show the results from data sets D_{M_1}, D_{M_2} and D_{M_3}, respectively. The figures on the 1^{st}, 2^{nd} and 3^{rd} columns show the results of alternative models M_1, M_2 and M_3, respectively. In each figure, for each temperature τ_n, the row n shows $\hat{R}$ of θ_m in the column m, for $n = 1, 2, ..., 10$ and $m = 1, 2, ...d$.

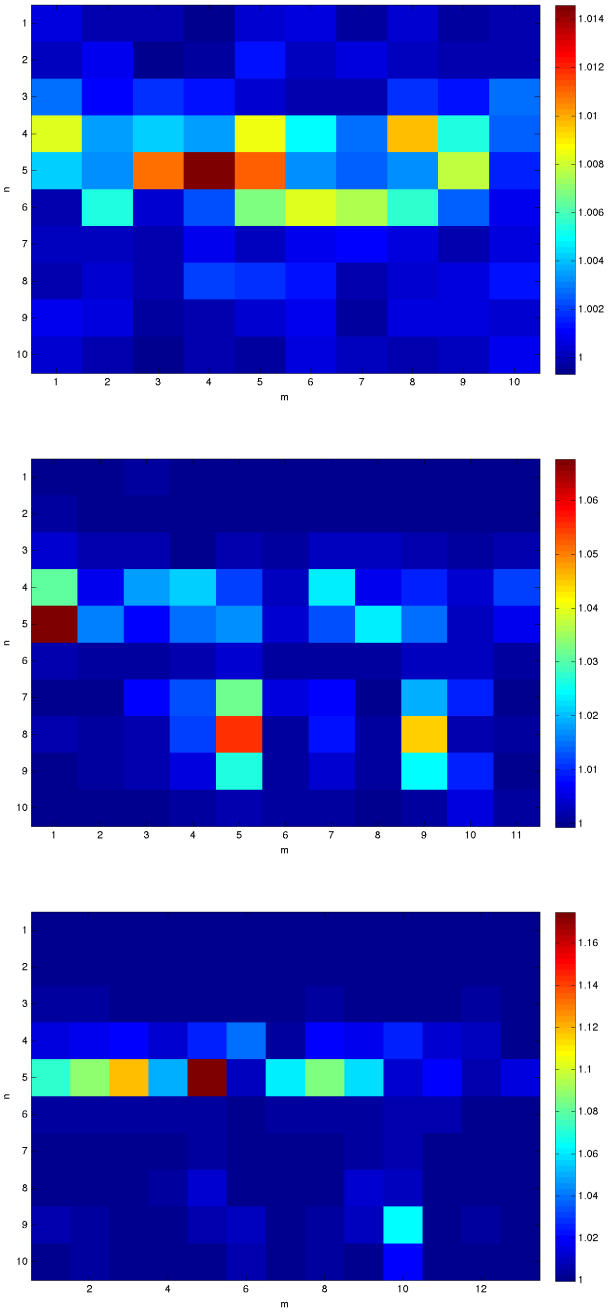

Figure 9.2: All the estimated potential scale reduction $\hat{R}$ of the sampling results for M_3 from the experimental data D_E are close to 1. The 1^{st}, 2^{nd} and 3^{rd} figures show the results of alternative models M_1, M_2 and M_3, respectively. For each temperature τ_n, the row n shows $\hat{R}$ of θ_m in the column m, for $n = 1, 2, ..., 10$ and $m = 1, 2, ...13$.

9.2.2 Estimated Parameter Posterior Distributions

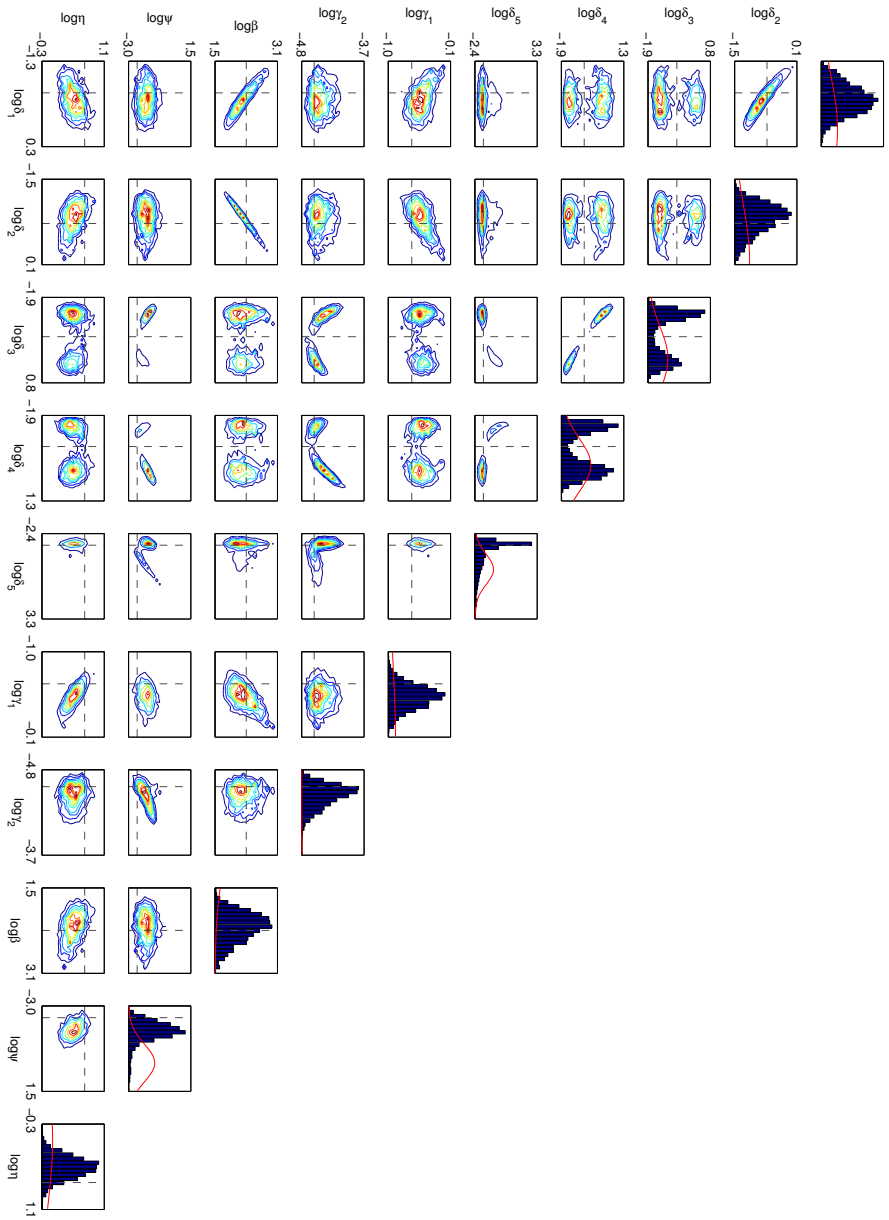

Figure 9.3: The estimated parameter posterior distribution for M_1 given the simulated data D'_{M_1}. The diagonal plots show the marginal distributions with respect to the corresponding parameters and the rest of the plots shows the dependences between two parameters with the dashed lines representing the true values. The prior distributions are plotted by solid red lines on the marginal distributions.

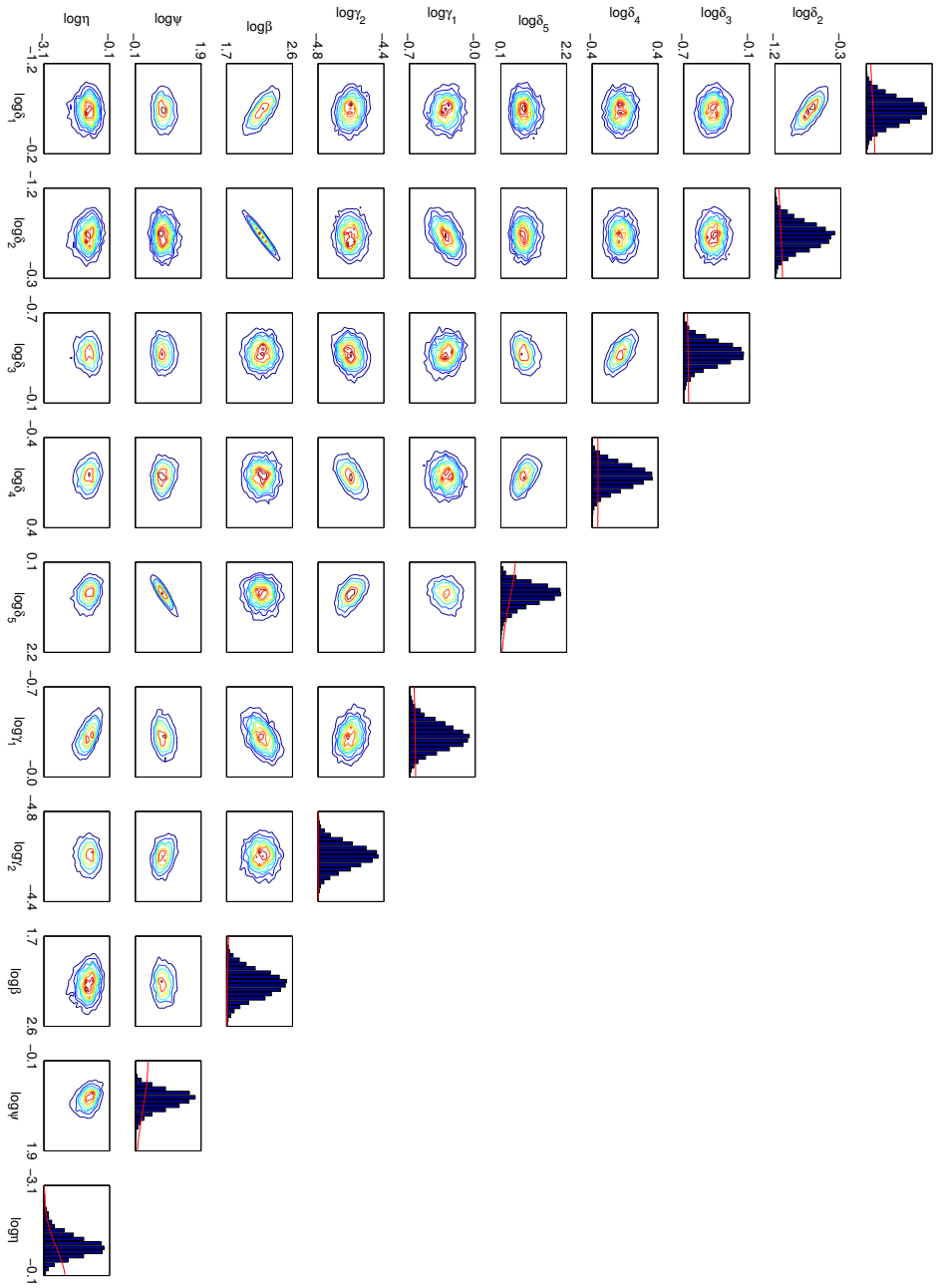

Figure 9.4: The estimated parameter posterior distribution for M_1 given the experimental data D_E. The diagonal plots show the marginal distributions with respect to the corresponding parameters and the rest of the plots shows the dependences between two parameters. The prior distributions are plotted by solid red lines on the marginal distributions.

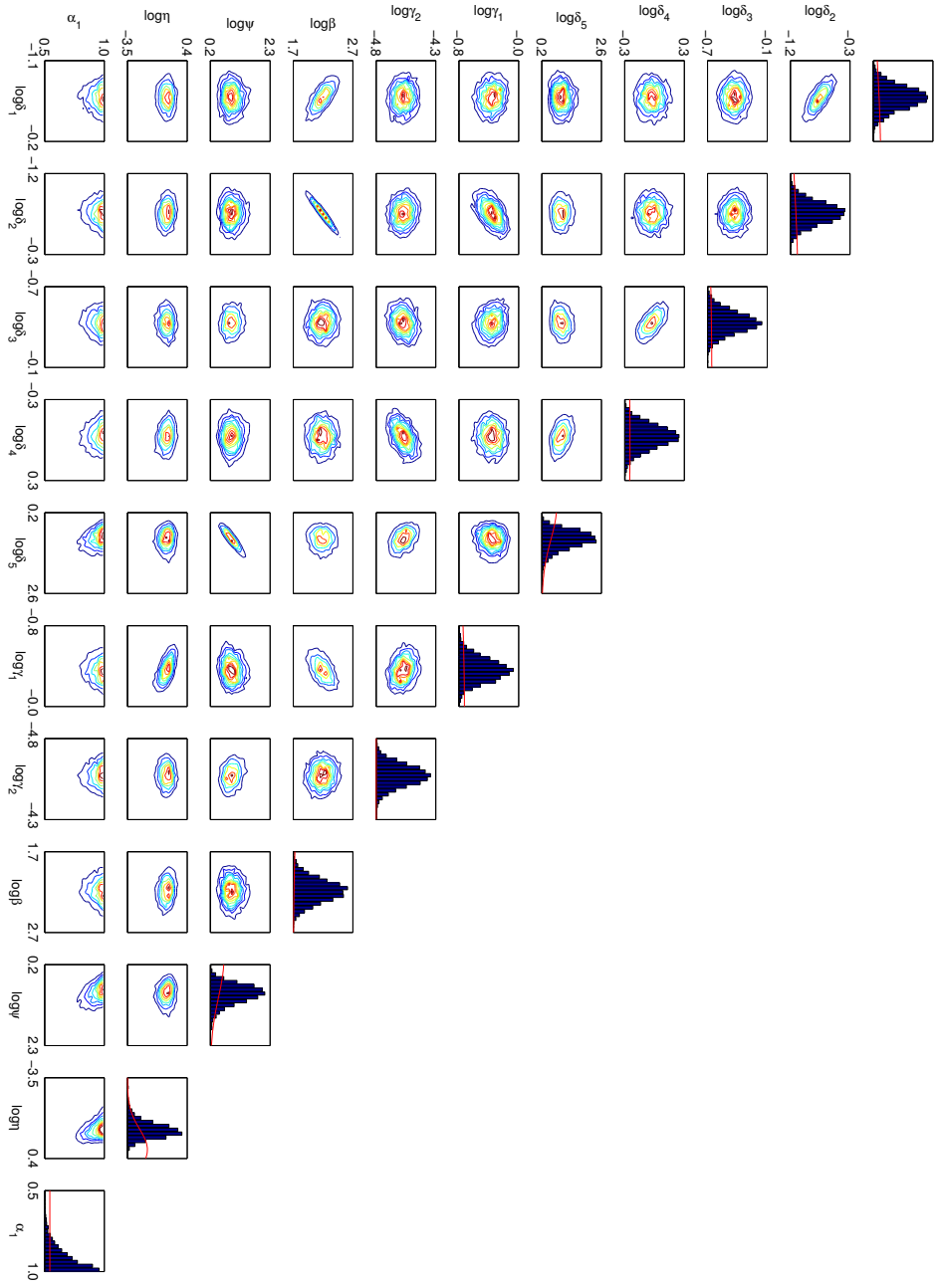

Figure 9.5: The estimated parameter posterior distribution for M_2 given the experimental data D_E. The diagonal plots show the marginal distributions with respect to the corresponding parameters and the rest of the plots shows the dependences between two parameters. The prior distributions are plotted by solid red lines on the marginal distributions.

Figure 9.6: The estimated parameter posterior distribution for M_3 given the experimental data D_E. The diagonal plots show the marginal distributions with respect to the corresponding parameters and the rest of the plots shows the dependences between two parameters. The prior distributions are plotted by solid red lines on the marginal distributions.

9.2.3 Posterior Predictive Distributions

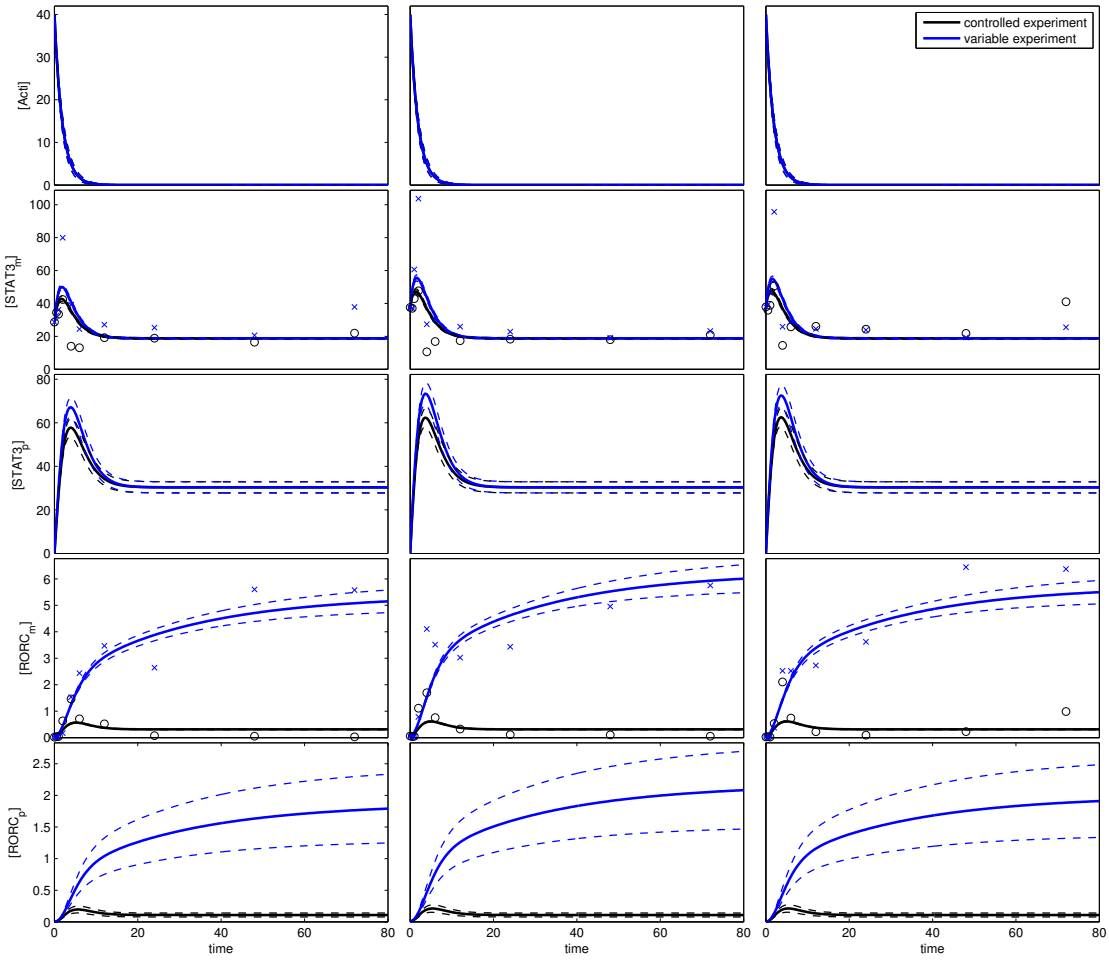

Figure 9.7: Predictive distributions generated using the dominant model, M_3. The figures on each row show time-dependent marginal predictive distributions of $[Acti]$, $[STAT3_m]$, $[STAT3_p]$, $[RORC_m]$ and $[RORC_p]$ respectively, using the mean (solid lines) and one standard deviation interval (dashed lines) that the measurement uncertainties are not included but parameter uncertainties only. The lines are plotted using black and blue color for controlled and variable experiments, respectively. The normalized RPKM data are plotted using black circles and blue cross marks for controlled and variable experiments, respectively. Both predictions and data are replicate-dependent, shown in the 1^{st}, 2^{nd} and 3^{rd} columns representing 1^{st}, 2^{nd} and 3^{rd} replicates, respectively.

Bibliography

[1] Jinfang Zhu, Hidehiro Yamane, and William E Paul. Differentiation of effector CD4 T cell population. *Annual review of immunology*, 28:445, 2010.

[2] Thomas Korn, Estelle Bettelli, Mohamed Oukka, and Vijay K Kuchroo. IL-17 and Th17 Cells. *Annual review of immunology*, 27:485–517, 2009.

[3] Casey T Weaver, Charles O Elson, Lynette A Fouser, and Jay K Kolls. The Th17 pathway and inflammatory diseases of the intestines, lungs and skin. *Annual review of pathology*, 8:477, 2013.

[4] Pawel Muranski and Nicholas P Restifo. Essentials of Th17 cell commitment and plasticity. *Blood*, 121(13):2402–2414, 2013.

[5] Edda G Schulz, Luca Mariani, Andreas Radbruch, and Thomas Höfer. Sequential polarization and imprinting of type 1 T helper lymphocytes by interferon-γ and interleukin-12. *Immunity*, 30(5):673–683, 2009.

[6] Andrew Yates, Robin Callard, and Jaroslav Stark. Combining cytokine signalling with T-bet and GATA-3 regulation in Th1 and Th2 differentiation: a model for cellular decision-making. *Journal of theoretical biology*, 231(2):181–196, 2004.

[7] Thomas Höfer, Holger Nathansen, Max Löhning, Andreas Radbruch, and Reinhart Heinrich. GATA-3 transcriptional imprinting in Th2 lymphocytes: a mathematical model. *Proceedings of the National Academy of Sciences*, 99(14):9364–9368, 2002.

[8] Tian Hong, Jianhua Xing, Liwu Li, and John J Tyson. A mathematical model for the reciprocal differentiation of T helper 17 cells and induced regulatory T cells. *PLoS computational biology*, 7(7):e1002122, 2011.

[9] Tian Hong, Jianhua Xing, Liwu Li, and John J Tyson. A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. *BMC systems biology*, 6(1):66, 2012.

[10] Luis Mendoza. A network model for the control of the differentiation process in Th cells. *Biosystems*, 84(2):101–114, 2006.

[11] Luis Mendoza and Fátima Pardo. A robust model to describe the differentiation of T-helper cells. *Theory in biosciences*, 129(4):283–293, 2010.

[12] Jukka Intosalmi, Helena Ahlfors, Sini Rautio, Henrik Mannerstöm, Zhi J Chen, Riitta Lahesmaa, Brigitta Stockinger, and Harri Lähdesmäki. Analyzing th17 cell differentiation dynamics using a novel integrative modeling framework for time-course rna sequencing data. *BMC systems biology*, 9(1):81, 2015.

[13] Mark D Robinson and Gordon K Smyth. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9(2):321–332, 2008.

[14] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.

[15] Mark D Robinson, Alicia Oshlack, et al. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3):R25, 2010.

[16] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biol*, 11(10):R106, 2010.

[17] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[18] Daniel Schmidl, Sabine Hug, Wei B Li, Matthias B Greiter, and Fabian J Theis. Bayesian model selection validates a biokinetic model for zirconium processing in humans. *BMC systems biology*, 6(1):95, 2012.

[19] Nial Friel and Anthony N Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607, 2008.

[20] Ben Calderhead and Mark Girolami. Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*, 53(12):4028–4045, 2009.

[21] Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D Watson. *Molecular biology of the cell*. Garland Science, 2007.

[22] Günter P Wagner, Koryu Kin, and Vincent J Lynch. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131(4):281–285, 2012.

[23] Brian Ingalls. *Mathematical Modelling in Systems Biology: An Introduction*. 2012.

[24] Ting Chen, Hongyu L He, George M Church, et al. Modeling gene expression with differential equations. In *Pacific symposium on biocomputing*, volume 4, page 4, 1999.

[25] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.

[26] Eugene J Putzer. Avoiding the Jordan canonical form in the discussion of linear systems with constant coefficients. *American Mathematical Monthly*, pages 2–7, 1966.

[27] Nicholas J Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications*, 26(4):1179–1193, 2005.

[28] William E Boyce, Richard C DiPrima, and Charles W Haines. *Elementary differential equations and boundary value problems*, volume 9. Wiley New York, 1992.

[29] Vladislav Vyshemirsky and Mark A Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839, 2008.

[30] Mark Girolami. Bayesian inference for differential equations. *Theoretical Computer Science*, 408(1):4–16, 2008.

[31] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[32] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[33] Mark Steyvers. Computational Statistics with Matlab, 2010.

[34] Ajay Jasra, David A Stephens, and Christopher C Holmes. On population-based simulation for static inference. *Statistics and Computing*, 17(3):263–279, 2007.

[35] Ben Calderhead and Mark Girolami. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. *Interface Focus*, 1(6):821–835, 2011.

[36] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[37] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, pages 223–242, 2001.

[38] Jan Hasenauer, Christine Hasenauer, Tim Hucho, and Fabian J Theis. ODE Constrained Mixture Modelling: A Method for Unraveling Subpopulation Structures and Dynamics. *PLoS computational biology*, 10(7):e1003686, 2014.

[39] Andreas Raue, Marcel Schilling, Julie Bachmann, Andrew Matteson, Max Schelke, Daniel Kaschek, Sabine Hug, Clemens Kreutz, Brian D Harms, Fabian J Theis, et al. Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335, 2013.